

seq_documentation_block:
 artifact-warning sequence (translated ALU class F) - human
 C:Species: Homo sapiens (man)
 C:Date: 31-Mar-1992 #sequence_rev1sion 11-Aug-1995 #text_change 11-Aug-1995
 C:Accession: F40201
 R:Claverie, J.M.
 Personal communication, 1992
 A:Reference number: A40201
 A:Contents: conceptual translations of ALU-repeat sequences
 A:Accession: F40201
 A:Molecule type: DNA
 A:Residues: 1-673 <CL>
 R:Claverie, J.M.
 Genomics 12, 838-841, 1992
 A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
 A:Reference number: A40200; MUID:92241891
 A:Contents: annotation
 C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 in-frame stop codons are shown as 'X'.
 C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

alignment_scores:
 Quality: 289.50 Length: 347
 Ratio: 1.713 Gaps: 17
 Percent Similarity: 48.703 Percent Identity: 32.853

alignment_block:
 US-09-030-606-174/rev x F40201 ..

Align seg 1/1 to: F40201 from: 1 to: 673

```

1459 TTTTTTTTTTTTTT.....TTTGACAGAGCTTACTC 1425
      |||||||
345 PhepPhepPhepPhepCysPheValpHeValpHeGluIleGlyProCysSe 361
      |||||||
1424 TGTTCGCCAGCTGAGTATAGTGTGATCTCACTCACTCACTC 1375
      |:::|||||
361 RasphrProAlaGlyValGlnTrp.GlnValLeuAlaHisCysSerLeu 377
      |||||||
1374 TGCTCCCATTCAGCAATTCCTGCTCCAGCTCCCAAGTAGCTGG 1325
      |||||||
378 AsnLeuGlySerSerSerProAlaSerValSerArgValAlaG1 394
      |||||||
1324 GATTACAGCGCGCTCCACACATATCCAGCTACTTTGATTTTAGTAC 1275
      |||||||
394 YlleThrGlyMetGlyArgHisSerTrpLeuIleValpHeLeuIleG 411
      |||||||
1274 AGACAGATTCACCATTTTGGCCAGCTGCTTGAACCTCTTACCTCA 1225
      |||||||
411 LuThGlnPheHisHisValaspGlnAlaGlyLeuIleLeuThrSer 427
      |||||||
1224 AGTATCTGCTGCTGCTGCTCCCAAGTCTGGATTACAGCATGAGC 1175
      |||||||
428 SerAspLeuProSerTrpSerProIlyValLeuGly***GlnAla**A1 444
      |||||||
1174 CACCTGCGCCAGC.....CTTTGTAGAGCTGATTTTCCGTTTC 1131
      |||||||
444 ArThrProSer*****PheValpHeGlyIlyPhePhePhePheP 461
      |||||||
1130 TCTTCGTTTTCACCTCTAT.....GAATCGTGTCACTGTTCCCTCT 1087
      |||||||
461 hePheAlaLeuPheLeuPheLeuArg***AlaLeuAlaLeuThrProArg 477
      |||||||
1086 GGGTACACAGTTGACCTTGAAACAACCTGTTGAATGCACAGTCCAC 1037
      |||||||
478 .....LeuGlyCysSerG1 482
      |||||||
1036 TTATCTTGATTTTTCATATAACATCAGAAAAATTTTGGAGAGTT 987
      |||||||
482 yLysPheTrpLeuThrAlaAlaSerThrSerTrp..... 493

```

```

986 GCAACAATTTAAAAAATTCAGACGACGAGCTGTAGCTAGAAATATCA 937
      |||||||
494 .....ValGlnAlaIleLeuLeuProLeuSerProVal***LeuGly 507
      |||||||
936 AAAAATTCACAAGAGTATATCATGATGCATTAAGATGATGATTAAT 887
      |||||||
507 ..... 507
886 CGACTATTATGTTATGTAGAGCTCCCGTCAACAGTACGATTTCT 837
      |||||||
508 .....LeuGlnAlaTrpAlaAlaIleProGly***PheMetIlyPhe. 521
      |||||||
836 AGTCAGGTTTTTGGGAGAGCAAAAGTATATAA..... 804
      |||||||
522 *****ArgHisSerPheThrMetLeuIleArgLeuValSerAsnSer** 538
      |||||||
803 .....GGATTTCATTTGGGTGG.....A 785
      |||||||
538 *ProGlnValIleCysProProGlyLeuProIysCysTrpAspArgA 555
      |||||||
784 GGTACCG.....CCCTCAGGCCCATGCTGTGAGAGTCAACTGCAT 741
      |||||||
555 TgLuProProHisProAla*****LeuPheLeuGlyValpHePhe 571
      |||||||
740 TTTATGTGTGTGTGTTTTCGCTCCCATGCTGTTGTTCTGTCTCTCT 691
      |||||||
572 PhePhe.....PheLeu.....LeuCysPheCysPhe***As 582
      |||||||
690 TCTTTGCTTCTCTTTATTCCTGTGTTATTTCTGTTTCTCTCAGTT 641
      |||||||
582 PargProLeuLeu***HisProGlyTrpSerAlaValAlaLaserPheG1s 599
      |||||||
640 CTCCTCTC.....AGTGTCACTGTCTC 618
      |||||||
599 erLeuGlnProGlnProProGlyPheLysArgPheSerCysLeuCysLeu 615
      |||||||
617 CCTGTGTGTCTCTCATCTCTGATCTGCGCAT.....GC 583
      |||||||
616 ProCysSerTrpAspTyr.....ArgHisGlyProProPheLeuAl 629
      |||||||
582 GGCCCTGTGTGTC.....TGTGTCTCCCCC 558
      |||||||
629 AsnLeuCysIlePheAsnArgAspThrValSerPro 641

```

seq_name: p1r4:A40201

seq_documentation_block:
 artifact-warning sequence (translated ALU class A) - human
 C:Species: Homo sapiens (man)
 C:Date: 31-Mar-1992 #sequence_rev1sion 11-Aug-1995 #text_change 11-Aug-1995
 C:Accession: A40201
 R:Claverie, J.M.
 Personal communication, 1992
 A:Reference number: A40201
 A:Contents: conceptual translations of ALU-repeat sequences
 A:Accession: A40201
 A:Molecule type: DNA
 A:Residues: 1-627 <CL>
 R:Claverie, J.M.
 Genomics 12, 838-841, 1992
 A:Title: Identifying coding exons by similarity search: Alu-derived and other potenti
 A:Reference number: A40200; MUID:92241891
 A:Contents: annotation
 C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames o
 in-frame stop codons are shown as 'X'.
 C:Comment: Any significant similarity of a predicted protein sequence to a portion of

alignment_scores:
 Quality: 285.00 Length: 125
 Ratio: 3.476 Gaps: 4
 Percent Similarity: 65.600 Percent Identity: 56.000


```

127 ArgSerSerArgProAlaTrpProThrTrpGlnAsnProIleSerThrLy 143
1282 AAATACAAAGAACTAGCTGATATGATGGCAGCGCCTGTAATCCAGCTA 1331
||||| 1282 1331
143 SASrThLyAsn***LeuGlyMetValAlaProValAlaProAlaArgT 160
||||| 143 160
1332 CTTCGGAGCGCTGAGCAGAGAAATTCCTTGAATATGGAGGACAGAGTTG 1381
||||| 1332 1381
160 hrgLInLyAlaGluAlaGlyLInTrpPhLysProGlyArgTrpSerLeu 176
||||| 160 176
1382 AAGCAGATGTGATCAGACCACTATACCTCCAGCTGGGCAACAGAGTTAG 1431
||||| 1382 1431
177 Gln***AlaGluIleSerProLeuHisSerSerLeuGlyAspArgValAr 193
||||| 177 193
1432 ACTCTGCTCAAAAAA***** 1459
||||| 1432 1459
193 gLeuCysLeuLysLysLysLysLysLys 202
||||| 193 202

```

seq_name: p1r4:A40201

seq_documentation_block:

artifact-warning sequence (translated ALU class A) - human
 C:Species: Homo sapiens (man)
 C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 11-Aug-1995
 R:Claverie, J.M.
 C:Accession: A40201
 A:Reference number: A40201
 A:Personal communication, 1992
 A:Reference number: A40201
 A:Contents: conceptual translations of ALU-repeat sequences
 A:Accession: A40201
 A:Molecule type: DNA
 A:Residues: 1-627 <CIA>
 R:Claverie, J.M.
 Genomics 12, 838-841, 1992
 A:Title: Identifying coding exons by similarity search: Alu-derived and other potential
 A:Reference number: A40200; MID:92241891
 A:Contents: annotation
 C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
 In-frame stop codons are shown as 'X'.
 C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

alignment_scores:
 Quality: 247.50 Length: 98
 Ratio: 3.300 Gaps: 2
 Percent Similarity: 76.531 Percent Identity: 61.224

alignment_block:

US-09-030-606-174 x A40201 ..

Align seg 1/1 to: A40201 from: 1 to: 627

```

1166 GCGAGGCGTG...GCTCATGCTGTAAATCCAGCACTTTGGAGCGGAGGC 1212
||||| 1166 1212
1 GLyTrgAlaArgSerHisAlaCysAsnProSerThrLeuGlyGlyGlnG 17
||||| 1 17
1213 AGGAGATCACTTGAAGTCAAGAGTTCAGACAGCCCTGGCCAAATAGT 1262
||||| 1213 1262
17 yGlyArgGle..ThrArgSerGlyAspArgAsnProGlyTrpHisGly 33
||||| 17 33
1263 GAAATCCCTGTCTACTAAATACAAAGTATAGCTGATATGGTGGCAG 1312
||||| 1263 1312
34 GluThrProSerLeuLeuLysIleGlnLysIleSerArgAla***CysG 50
||||| 34 50
1313 GCGCCTGTAATCCAGCTTGGAGGCTGAGGAGGAGGAGGAGATTGCTTGA 1362
||||| 1313 1362
50 yArgLeu***SerGlnLeuGlyArgLeuArgGlnGlnLysnGlyValA 67
||||| 50 67
1363 ATATGGGAGGAGAGAGTGAAGTGAAGTGAATCAGACCACTATATCTCA 1412
||||| 1363 1412
67 snProGlyGlyGlyAlaCysSerGluProArgSerArgHisCysThrPro 83
||||| 67 83

```

```

1413 GCTGGGCAACAGAGTAAAGACTCTGTCTCAAAAAAAAAAAAA 1454
||||| 1413 1454
84 AlAeuAlaThrGluArgAspSerValSerGluLysAsnLys 97
||||| 84 97

```

seq_name: p1r2:A46010

seq_documentation_block:

X-linked retinopathy protein (C-terminal, clone XEH.8c) - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
 C:Accession: A46010
 R:Wong, P.; MacDonald, I.M.; Sood, R.; Smith, C.; Pilon, R.; Tenniswood, W.
 Genomics 15, 467-471, 1993
 A:Title: Identification and partial characterization of a candidate gene for X-linked
 A:Reference number: A46010; MID:93224131
 A:Accession: A46010
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-100 <MON>
 A:Cross-references: GB:S58722; NID:9299470; PID:9299471
 A:Note: sequence extracted from NCBI backbone (NCBIN:129339, NCBI:P:129340)

alignment_scores:
 Quality: 246.50 Length: 117
 Ratio: 3.160 Gaps: 4
 Percent Similarity: 66.667 Percent Identity: 51.282

alignment_block:

US-09-030-606-174/rev x A46010 ..

Align seg 1/1 to: A46010 from: 1 to: 100

```

1459 TTTT*****TTTGTGACAGAGCTTACTCTGTGCCCCAGCTGG 1410
||||| 1459 1410
1 PhePhePhePhePhePheGluThrGluSerCysSerValAlaGluAlaG 17
||||| 1 17
1409 AGTATAGTGTGATGTCACTCACTTCAACCTGCTGCCATATCA 1360
||||| 1409 1360
17 yValGlnTrpCysAspLeuGlySerLeuLysSerProPro.ProGlySer 33
||||| 17 33
1359 AGCAATTCCTCGCTCAGCGCTCCCAAGTAGCTGGATTACAGCGCGCTG 1310
||||| 1359 1310
34 SerAspSerProAlaSerAlaSerArgValAlaGlyLeuThrGlyMetH 50
||||| 34 50
1309 CCACCATATCCAGCTACTTTGTATTTTAGTACAGACGAGATTTCAC. 1261
||||| 1309 1261
50 SHShIShThrGlnLeuIlePheValPheLeuValGluThrGlySerHisM 67
||||| 50 67
1260 ..CATTTGGCAGAGCTGTGATGAACTCCTTACCTCACTCAAGTGAT..... 1219
||||| 1260 1219
67 etGlnLeuSerAspSerThrLeuValIleThrThrAlaGlnAsnAlaLys 83
||||| 67 83
1218 CTGCGCTCGCTCGCTCCCAAGTCTGGGATTACAGCATAGCACCACCT 1169
||||| 1218 1169
84 IleThrAlaTrgAlaProArgAspLeu..... 92
||||| 84 92
1168 GCCCAGGCTCTTTTGAAGCTTGATTTTTCCTGTTCTCTTGCTGTTT 1120
||||| 1168 1120
93 .....PhePhePhePhePhePhePhe 100
||||| 93 100

```

seq_name: p1r2:I38022

seq_documentation_block:

hypothetical protein - human
 C:Species: Homo sapiens (man)
 C:Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 01-Nov-1996
 C:Accession: I38022
 R:Yang, S.S.; Zhang, K.; Vieira, W.; Teub, J.V.; Zellstra-Ryalls, J.H.; Somerville, R.
 Cancer Res. 50, 5636-5667, 1990
 A:Title: A human hepatocellular carcinoma 3.0-kilobase DNA sequence transforms both r
 A:Reference number: I38021
 A:Accession: I38022

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-196 <RES>
 A:Cross-references: EMBL:X55777; NID:9288143; PID:9288145

Alignment_scores:
 Quality: 242.00 Length: 78
 Ratio: 3.841 Gaps: 0
 Percent Similarity: 80.769 Percent Identity: 69.231

Alignment_block:
 US-09-030-606-174 x J38022 ..

Align seg 1/1 to: I38022 from: 1 to: 196

```

1221 CACTGAGGTAGAGTTAAAGCCAGCTGGCCAAATGTGAATCT 1270
||||| ||||||| ||||||| ||||||| |||||||
3 HistLeuArgSerGlyValGlnAspTyrProGlyGlnHisGlyLysIlePr 19
1271 GTCTGTAATAAATACAAAGTTAGCTGATATGGTGGCGGCTGT 1320
||||| ||||||| ||||||| ||||||| |||||||
19 OSerLeuLeuysIleGlnGlnLeuHisGlyLysIlePr 36
1321 AATCCAGCTACTGGAGGCTGAGGAGAGAAATGCTGAATATGGA 1370
||||| ||||||| ||||||| ||||||| |||||||
36 InSerGlnLeuLeuArgArgLeuArgGlnGlnHisLeuAsnSerGly 52
1371 GGCAAGGTGAGTGAAGTGTGAGATCAGCCACTTACTCCAGCTGGGC 1420
||||| ||||||| ||||||| ||||||| |||||||
53 GlyArgGlyCysSerLeuProLysSerHisLeuCysIleProAlaTyrVa 69
1421 AACAGAGTAGACTGTCTCTCAAAAAAATAAAAA 1454
||||| ||||||| ||||||| ||||||| |||||||
69 IThrGlnGlyAspSerValSerLysGlnAsnLys 80

```

seq_name: p1r2:J01472
 seq_documentation_block:
 trypsin (EC 3.4.21.4) V precursor, b-form - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 08-Sep-1997
 C:Accession: J01472
 R:Kang, J.; Wiegand, U.; Mueller-Hill, B.
 Gene 110, 181-187, 1992
 A:Title: Identification of cDNAs encoding two novel rat pancreatic serine proteases.
 A:Reference number: J01471; M01D:92165037
 A:Accession: J01472
 A:Molecule type: mRNA
 A:Residues: 1-246 <KAN>
 A:Cross-references: EMBL:X59013; NID:957414; PID:957415
 A:Experimental source: pancreas
 C:Superfamily: trypsin; trypsin homology
 C:Keywords: hydrolase; protein digestion; serine proteinase; zymogen
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-24/Domain: activation peptide #status predicted <ACT>
 F:25-246/Product: trypsin V, b-form #status predicted <MNT>
 F:25-239/Domain: trypsin homology <TRY>
 F:31-160,49-65,133-233,140-206,171-185/Disulfide bonds: #status predicted
 F:64,108,200/Active site: His, Asp, Ser #status predicted

Alignment_scores:
 Quality: 237.50 Length: 209
 Ratio: 2.220 Gaps: 6
 Percent Similarity: 51.196 Percent Identity: 31.100

Alignment_block:
 US-09-030-606-174 x J01472 ..

Align seg 1/1 to: J01472 from: 1 to: 246

3 TCAGCGCAGACTGTTCCAGAAAGTAGTGCAGAGACTCTACACCATGG 52

```

||||| ||||||| ||||||| ||||||| |||||||
61 SerAlaAlaHisCysTyrHisProGlnLeuGln.....ValAr 73
53 GCTGGGCGCTGCACAGTCTTGAGCGCGACCAAGAGCGGAGCGAATGG 102
||||| ||||||| ||||||| ||||||| |||||||
73 GLeuGlyGlnHisAsnIle...TyrGlnIleGlnGlyAlaGlnGlnPheI 89
103 TGGAGCGCCAGCTCTCCGTACGGACCCAGAGTACCAAGACCCCTGGTC 152
||||| ||||||| ||||||| ||||||| |||||||
89 LeAspAlaAlaLysMetIleLeuHisProAspTyrAspLysTyrThrVal 105
153 GCTAACGACTCATGCTCATCAAGTTGAGCAGAAATCCGTCCGAGTCTGA 202
||||| ||||||| ||||||| ||||||| |||||||
106 AspAsnAspIleMetLeuIleLysLeuLysSerProAlaThrLeuAsnSe 122
203 CACCATCCGAGCATCAGCATTCGCTGCGACATGCCCTACCCGGGAACT 252
||||| ||||||| ||||||| ||||||| |||||||
122 LysValSerThrIleProLeuProGlnTyrCysProThrIleGlyThrG 139
253 CTGCGCTGCTTCTGCGTGGGCTGCTGCGGAGGAGGAGACTCAGCGGT 302
||||| ||||||| ||||||| ||||||| |||||||
139 LysCysLeuValSerGlyTyrProLysValLeuLysPheGly..... 151
303 GTGTCTGTGCCCTTTCAGAGAGGTCTCTGCCAGTCCGCGGGGCTGAC 352
||||| ||||||| ||||||| ||||||| |||||||
151 ..... 151
353 CCAGAGCTCTGCTGCCAGCAGAAATG..CTTACCGTGTCTGATGCTGA 401
||||| ||||||| ||||||| ||||||| |||||||
152 ..... 162
402 ACGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 451
||||| ||||||| ||||||| ||||||| |||||||
162 SpAlaProValLeuSerAspSerValCysHisLysAlaTyrProArgGln 178
452 TACCAACCCCAACATGTTCTGCGCGCGCGGAGGAGCAAGCAGAAAGATC 501
||||| ||||||| ||||||| ||||||| |||||||
179 IleThrAsnAsnMetPheCysLeuGlyPheLeuGlnGlyLysAspSe 195
502 CTGCACGCTGAGAGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGT 551
||||| ||||||| ||||||| ||||||| |||||||
195 rCysGlnTyrAspSerGly..... 201
552 GGAGAGGGGGAGAGACAGACACAGGCGCG.....CATGGCAG 592
||||| ||||||| ||||||| ||||||| |||||||
202 ..... 209
593 ATGCAG.....AGATGAGAGAGAC 610
||||| ||||||| ||||||| ||||||| |||||||
210 ValGlnGlyIleValSerTyrProLysP 218

```

seq_name: p1r2:J01471
 seq_documentation_block:
 trypsin (EC 3.4.21.4) V precursor, a-form - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 08-Sep-1997
 C:Accession: J01471; M01D:92165037
 R:Kang, J.; Wiegand, U.; Mueller-Hill, B.
 Gene 110, 181-187, 1992
 A:Title: Identification of cDNAs encoding two novel rat pancreatic serine proteases.
 A:Reference number: J01471; M01D:92165037
 A:Accession: J01471
 A:Molecule type: mRNA
 A:Residues: 1-246 <KAN>
 A:Cross-references: EMBL:X59012; NID:957412; PID:957413
 A:Experimental source: pancreas
 C:Superfamily: trypsin; trypsin homology
 C:Keywords: hydrolase; protein digestion; serine proteinase; zymogen
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-24/Domain: activation peptide #status predicted <ACT>
 F:25-246/Product: trypsin V, a-form #status predicted <MNT>
 F:25-239/Domain: trypsin homology <TRY>

F:31-160,49-65,133-233,140-206,171-185/Disulfide bonds: #status predicted
F:64,108,200/Active site: His, Asp, Ser #status predicted

alignment_scores: Quality: 237.50 Length: 209
 Ratio: 2.220 Gaps: 6
Percent Similarity: 51.196 Percent Identity: 31.100

alignment_block:

US-09-030-606-174 x JQ1471 ..

Align seg 1/1 to: JQ1471 from: 1 to: 246

```
3 TCAGCCGACACTGTTCCAGAAAGTAGTCAGAGCTCCTACACATCGG 52
|||||
61 SerAlaAlaHisCysTyrHisProGlnLeuGln.....ValAr 73
53 GCTGGGCTGTCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGG 102
|||||
73 GLeuGlyGlnHisAsnIle...TyrGluIleGlyGlnGlnGlnPheI 89
103 TGGAGGCCAGCTCTCCGTAAGGAGCCAGAGTACAAACAGACCTTGCTC 152
|||||
89 IeAspAlaAlaLysMetIleLeuHisProAspTyrAspLysTrpThrVal 105
153 GCTAAGACCTCATGCTCATCAAGTTGGAGAAATCCGTGCCAGTCTGA 202
|||||
106 AspAsnAspIleMetLeuIleLysLeuLysSerProAlaThrLeuAsn 122
203 CACCATCCGAGCATCAGCATTCGTCAGTGCCTCCACCGCGGGAGACT 252
|||||
122 LysValSerThrIleProLeuProGlnTyrCysProThrAlaGlyThrg 139
253 CTTCGCTCGTTTCTGGCTGGGCTCTGCTGCGAAGCGTACGTCACGGGT 302
|||||
139 LncLysLeuValSerGlyTyrGlyValLeuLysPheGly..... 151
303 GTCTGTCTGCCCTCTCAAGAGGTCTCTGCCAGTCGGGGGGCTGAC 352
|||||
151 ..... 151
353 CCAGAGCTCTGCTCCAGGAGAGATG.CCTACCGTCTGCTGCTGTA 401
|||||
152 .....PheGlnSerProSerValLeuGlnCysLeuA 152
402 ACCTGTCGCTGCTGTCTGANGAGTCTGCANTTAAGCTTAAGACCGCTG 451
|||||
162 sPaIaProValLeuSerAspSerValCysHisLysAlaTyrProArgGln 178
452 TACACCCCANCATGTTCTGCGCGGGGAGGAGGACCAAGCAAGAGATC 501
|||||
179 lIeThraSnmMetPheCysLeuGlyPheLeuGlnGlyLysAspSe 195
502 CTGCAACGTGAGAGAGGGAAGGGAGGCGACGACTCAGGAAGGAT 551
|||||
195 rCysGlnTyrAspSerGly..... 201
552 GGAAGAGGGGAGAGACAGACACAGGCGG.....CATGGCAG 592
|||||
202 .....GlyProValValCysAsnGlyGln 209
593 ATGCAG.....AGATGGAGAGAC 610
|||||
210 ValGlnGlyIleValSerTrpGlyAsp 218
```

seq_name: p1r1:TRDGC

seq_documentation_block:

trypsin (EC 3.4.21.4) precursor, cationic - dog
N:Alternate names: cationic trypsinogen
C:Species: Canis lupus familiaris (dog)
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 22-Jan-1999

C:Accession: B26273
R:Pinsky, S.D.; LaForge, K.S.; Scheele, G.
Mol. Cell. Biol. 5, 2669-2676, 1985
A:Title: Differential regulation of trypsinogen mRNA translation: full-length mRNA se
A:Reference number: A26273; MUID:86284628
A:Accession: B26273
A:Molecule type: mRNA
A:Residues: 1-246 <PIN>

A:Cross-references: GB:M1590; NID:g164096; PID:g164097
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-23/Domain: activation peptide #status predicted <APT>
F:24-246/Product: trypsin, cationic #status predicted <ENT>
F:24-239/Domain: trypsin homology <TR>
F:30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted
F:63,107,200/Active site: His, Asp, Ser #status predicted
F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Gln) #status predicted

alignment_scores: Quality: 236.00 Length: 202
 Ratio: 2.107 Gaps: 7
Percent Similarity: 55.446 Percent Identity: 30.693

alignment_block:

US-09-030-606-174 x TRDGC ..

Align seg 1/1 to: TRDGC from: 1 to: 246

```
3 TCAGCCGACACTGTTCCAGAAAGTAGTCAGAGCTCCTACACATCGG 52
|||||
60 SerAlaAlaHisCysTyrLysSerArgIleGln.....ValAr 72
53 GCTGGGCTGTCACAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGG 102
|||||
72 GLeuGlyGlnTyrAsnIle...AlaValSerGlnGlyGlnGlnPheI 88
103 TGGAGGCCAGCTCTCTCCAGGAGCCAGAGTACAAACAGACCTTGCTC 152
|||||
88 IeAsnAlaAlaLysIleIleArgHisProArgTyrAsnAlaAsnThrIle 104
153 GCTAAGACCTCATGCTCATCAAGTTGGAGAAATCCGTGCCAGTCTGA 202
|||||
105 AspAsnAspIleMetLeuIleLysLeuLysSerProAlaThrLeuAsn 121
203 CACCATCCGAGCATCAGCATTCGTCAGTGCCTCCACCGCGGGAGACT 252
|||||
121 rArgValSerAlaIleAlaLeuProLysSerCysProAlaAlaGlyThrg 138
253 CTTCGCTCGTTTCTGGCTGGGCTCTGCTGCGAAGCGTACGTCACGGGT 302
|||||
138 LncLysLeuIleSerGlyTyrGly..... 145
303 GTCTGTCTGCCCTCTCAAGAGGTCTCTGCCAGTCCGGGGGGCTGAC 352
|||||
146 .....AsnThrGlnSerIleIle..... 151
353 CCAGAGCTCTGCTCCAGGAGAGATG.CCTACCGTCTGCTGCTGTA 402
|||||
152 .....Gln.AsnTyrProAspValLeuGlnCysLeuLys 162
403 CGTGTGCTGCTGTCTGANGAGTCTGCANTTAAGCTTAAGACCGCTGT 452
|||||
162 sAlaProIleLeuSerAspSerValCysArgAsnAlaTyrProGlyGlnI 179
453 ACCACCCCANCATGTTCTGCGCGGGGAGGAGGACCAAGCAAGAGATC 502
|||||
179 IeSerSerAsnMetCysLeuGlyTyrMetClnGlyLysAspSer 195
503 TGCACGTGAGAGAGGGAAGGGAGGCGACGACTCAGGGAAGGAGTG 552
|||||
196 CysGlnGlyAspSerGly..... 201
```

553 GAGAGGGGAGAGACAGACACAGGCGG.....CATGGCGAGA 593
||||| :|||||
202GlyProValValcysnnglylul 210
594 TCAG 598
:||||
210 eugln 211

seq_name: p1r4:E40201

seq_documentation_block:

artifact-warning sequence (translated ALU class E) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 11-Aug-1995
C:Accession: E40201
R:Claverie, J.M.
Personal communication, 1992
A:Reference number: A40201
A:Contents: conceptual translations of ALU-repeat sequences
A:Accession: E40201
A:Molecule type: DNA
A:Residues: 1-597 <CLAE>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: ALU-derived and other potential
A:Reference number: A40200; MUID:92241891
A:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of C
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

alignment_scores:
Quality: 233.50 Length: 103
Ratio: 3.243 Gaps: 7
Percent Similarity: 69.903 Percent Identity: 56.311

alignment_block:

US-09-030-606-174/rev x E40201 ..

Align seg 1/1 to: E40201 from: 1 to: 597

1450 TTTTGTGAGACAGACTTACT...GTGCCCCAGCTGAGATA 1404
||||| :|||||
501 PhePhePheLeuArgArgSerPheAlaLeuValAlaGlnAlaGlyValAr 517
||| :|||||
1403 GTGGTGTGATCACTCA.....CTTCAACCTGCGCTC 1369
||| :|||||
517 gTPrHismsnleuthralaasnPhelalSerTrpValGlnAlaIle.... 532
||| :|||||
1368 CCAATATTCAGCAATCTCTGCTGCTGAGCTCCCAAGTAGTGGATTAC 1319
:|||||
533LeuSerCysleuSerleuProSerSerTrpAspIyr 544
||| :|||||
1318 AGGGCGCCCGCACCATATCCAGCTACTTTGTATTATTATTAGTAGAGAG 1269
||||| :|||||
545 ArgHisAlaProProAlaArgProAlaasn.PheIlePheLeuValGlnMetg 561
||||| :|||||
1268 GATTTCACCATTTTGGCCAGGCTGCTTGAACCTTACCTCAAGTAT 1219
||||| :|||||
561 LypheLeuHisValGlyGlnAlaGlyLeuLysLeuProThrSerIlyasp 577
||||| :|||||
1218 CTGCTGCGCTGCG.....CTCCCAAGTGTGGATTACAGGCGATGAGC 1175
||||| :|||||
578 ...Pro.ProlArgLeuGlnLeuProLysArgTrpAspIyrAlaGlnGluL 593
||| :|||||
1174 CAGCC 1170
|||
593 eupro 594
seq_name: p1r2:S05494

seq_documentation_block:

trypsin (EC 3.4.21.4) IV precursor - rat
N:Alternate names: 23k protein; trypsinogen IV precursor
C:Species: Rattus norvegicus (Norway rat)
C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 08-Sep-1997
C:Accession: S05494
R:Luetcke, H.; Rausch, U.; Vasiloudes, P.; Scheele, G.A.; Kern, H.F.
Nucleic Acids Res. 17, 6736, 1989
A:Title: A fourth trypsinogen (P23) in the rat pancreas induced by CCK.
A:Reference number: S05494; MUID:89386010
A:Accession: S05494
A:Molecule type: mRNA
A:Residues: 1-247 <LNE>
A:Cross-references: EMBL:J15679; NID:956813; PID:956814
C:Superfamily: trypsin; trypsin homology
C:Keywords: calcium binding; hydrolase; protein digestion; serine proteinase; zymogen
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-23/Domain: activation peptide #status predicted <APT>
F:24-247/Product: trypsin IV #status predicted <MAT>
F:24-240/Domain: trypsin homology <TRY>
F:30-161,49-65,133-234,140-207,172-186/Disulfide bonds: #status predicted
F:54,108,201/Active site: His, Asp, Ser #status predicted
F:76,78,81,86/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

alignment_scores:
Quality: 233.00 Length: 173
Ratio: 2.354 Gaps: 4
Percent Similarity: 57.225 Percent Identity: 32.370

alignment_block:

US-09-030-606-174 x S05494 ..

Align seg 1/1 to: S05494 from: 1 to: 247

3 TCAGCCGACAGCTGTTCCAGAAAGTAGAGAGAGCTCTACACCATCGG 52
||||| :|||||
61 SerAlaAlaHisCysTyrLysArgLysLeuGln.....ValAr 73
||| :|||||
53 GCTGGGCTGCACAGTCTTGAGCGCGACCAAGAGCCAGGAGCCAGATGG 102
||||| :|||||
73 gLeuGlyGlnHismsnIleHisVal..LeuGlnGlyGlyGlnPheI 89
:|||||
103 TGAGAGCCAGCTCTCGCTACGGCAGCCAGATACAAAGACCTTCTCTC 132
:|||||
89 leaSpAlaGlnLysIleIleArgHisProGlnIlyrAsnLysAspThrLeu 105
||||| :|||||
153 GCTAACGACCTCATGCTCATCAAGTTGGAGACAAATCCGTCGCGAGTGA 202
||||| :|||||
106 AspAsnAspIleMetLeuIleLysLeuLysSerProAlaValLeuAsnSe 122
||| :|||||
203 CACCATCCGGAGCATACGATGCTTGCAGTGCCTACCGCGGGAGACT 252
||| :|||||
122 rGlnValSerThrValSerLeuProArgSerCysAlaSerThrAspAlaG 139
||| :|||||
253 CTGCTCGTGTCTGCTGGGGGTCTGCTGGCGAAGGAGTACGCGGT 302
:|||||
139 IncLysLeuValSerGlyTrpGly..... 146
||| :|||||
303 GTGTGTCTGCGCTTTCAAGAGAGTCTCTGCCAGTCGGGGGCGTGCAC 352
||| :|||||
147As 147
||| :|||||
353 CCAAGAGCTGCGTCCAGGACAGATGCTACCGTGCAGTGCAGTGA 402
||| :|||||
147 nThrValSer..IleGlyGlyLysTyrProAlaLeuLeuIncLysLeuG 163
||| :|||||
403 CGTGTGCGTGTCTGAGAGAGTGTGCANTAAAGCTGTATGACCCGCTGT 452
||| :|||||
163 uAlaProValLeuSerAlaSerSerCysLysSerTyrProGlyGlnI 180
||| :|||||
453 ACACCCCAACATCTTCTGCGCGCGGAGGAGGCAAGACCAAGAGACTCC 502

```

||||||| ||| ::| |||||||
180 lethrserasmetphecysleuglypheleuglyglylysaspser 196
503 TGCACGCTGAGAGGGG 520
||||| ::|||
197 Cysaspolyaspsery 202

```

seq_name: p1r1:TRBOTR

seq_documentation_block:

```

trypsin (EC 3.4.21.4) precursor - bovine
N:Contains: trypsinogen
C:Species: Bos primigenius taurus (cattle)
C:Date: 24-Apr-1984 #sequence_revision 28-Feb-1986 #text_change 18-Jul-1997
C:Accession: A00164; A00946; S08774
R:Miles, O.; Holeysovsky, V.; Tomasek, V.; Sorm, F.
Biochem. Biophys. Res. Commun. 24, 346-352, 1966
A:Title: Covalent structure of bovine trypsinogen. The position of the remaining amides.
A:Reference number: A00164; MUID:67168848
A:Accession: A00164
A:Molecule type: protein
R:Residues: 1-57, 'Q', 59-67, 'Q', 69-150, 'N', 152-176, 'N', 178-229 <MIK>
R:Hartley, B.S.
Philos. Trans. R. Soc. Lond. B257, 77-87, 1970
A:Reference number: A93755
A:Contents: annotation; revisions
R:Tittani, K.; Ericsson, L.H.; Neurath, H.; Walsh, K.A.
Biochemistry 14, 1358-1366, 1975
A:Title: Amino acid sequence of dogfish trypsin.
A:Reference number: A00950; MUID:75146445
A:Contents: annotation; revisions
A>Note: the sequence agrees with that shown
R:Bode, W.; Schwager, P.
J. Mol. Biol. 98, 693-717, 1975
A:Title: The refined crystal structure of bovine beta-trypsin at 1.8 angstrom resolution
A:Reference number: A92934; MUID:76072097
A:Contents: annotation; X-ray crystallography; binding sites for calcium, substrate, and
C:Comment: Trypsinogen is synthesized in the acinar cells of the pancreas.
C:Comment: Autocatalytic cleavage after Lys-6 leads to beta-trypsin by releasing a termi
s pseudotrypsin. A cleavage may also occur after Arg-105.
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen
F:1-229/Product: trypsinogen #status experimental <ZYM>
F:1-6/Domain: activation peptide #status experimental <APT>
F:7-222/Domain: trypsin homology <TRY>
F:7-131,132-229/Product: alpha-trypsin #status experimental <MPT>
F:6-7/Cleavage site: Lys-116 (enteropeptidase) #status experimental
F:13-143,31-47,115-216,122-189,154-168,179-203/Disulfide Bonds: #status experimental
F:46,90,183/Active site: His, Asp, Ser #status experimental
F:58,60,63,68/Binding site: calcium (Glu, Asn, Val, Glu) #status experimental
F:131-133/Cleavage site: Lys-Ser (autolytic) #status experimental

```

alignment_scores:

```

Quality: 226.50 Length: 184
Ratio: 2.137 Gaps: 6
Percent Similarity: 57.609 Percent Identity: 33.152

```

alignment_block:

US-09-030-606-174 x TRBOTR ..

Align seg 1/1 to: TRBOTR from: 1 to: 229

```

3 TCAGCCGACACTGTTCCAGAGTGAGTCACAGGCTCTACACATCGG 52
||||| ::|||
43 SerAlaIahiscystylrlysseryIleGln.....ValAr 55
53 GCTGGGCTCGACAGCTTGAGCGCCGACAGAGCCAGGAGGAGATGG 102
||||| ::|||
55 gIeuglygluAspAsnIleAsnVal..ValGIuGlyAsnGIuGlnPheI 71
103 TGAGGCGACGCTCTCGTAGGCGACCCAGAGTAGACAGACACCTTGCTC 152
:::||||| ::|||

```

```

71 leSerAlaSerlysserIleValHisProSerTyrAsnSerAsnThrIeu 87
153 GCTACGACCTCATGCTCATCAAGTTGACAGAAATCCGTGTCGAGTCTGA 202
||||| ::|||
88 AsnAspIleleuIleleuIlelysserIleValHisProSerTyrAsn 104
203 CACATCCGAGCATGAGCATGCTTCGAGTGGCCGAGTCCGCGGGGAGACT 252
: ::| ||||| ::|||
104 rArgValAlaSerIleSerIleuProThrSerCysAlaSerAlaGlyThr 121
253 CTGCGCTGTTCTGCGTGGGCTGCTGCGACGAGGAGCTCAGCGGT 302
::| ::|||
121 IncysleuIleSerGlyTrpGly.....AsnThrlysserSery 134
303 GGTGTCCTCCCTCTTCAAGAGGAGTCTTCCCACTGCGGGGGGCTGAC 352
::| ::|||
135 ..ThrSer.....
353 CCAGAGCTCTGCGTCCAGCAGAGCAATGCCCTACCGTGTGTCAGTGGTGA 402
||| ||||| ::|||
137 .....TyrProAspValIleuIlyCysleuIly 145
403 CGTGTGCTGTGTCTGANGAGTCTGCANTAGCTCATGACCCGCTGT 452
::| ::|||
145 sAlaProIleleuSerAspSerSeryCysIleSerAlaTyrProGlyGlnI 162
453 ACCACCCCAACATGTTCTGCGCGCGCGGAGGAGGACACAGAGACATCC 502
||| ::|||
162 lethrserasmetphecysalaglytyrleuglyglylysaspser 178
503 TGCAC.....GTGACAGAGAGGAGGAAGGAGGAGG 531
||| ::|||
179 CysGlnGlyAspseryglylyProValValCysSeryglylyleuglnI 195
532 C 532
195 Y 195

```

seq_name: p1r2:A27547

seq_documentation_block:

```

trypsin (EC 3.4.21.4) precursor, cationic - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Jun-1998 #sequence_revision 30-Jun-1998 #text_change 20-Mar-1998
C:Accession: A27547
R:Fletcher, T.S.; Alhadeff, M.; Craik, C.S.; Largman, C.
Biochemistry 26, 3081-3086, 1987
A:Title: Isolation and characterization of a cDNA encoding rat cationic trypsinogen.
A:Reference number: A27547; MUID:87271609
A:Accession: A27547
A:Molecule type: mRNA
A:Residues: 1-247 <FLN>
A:Cross-references: GB:M16624; NID:g206498; PID:g206499
C:Superfamily: trypsin; trypsin homology
C:Keywords: calcium binding; hydrolase; protein digestion; serine proteinase
F:25-240/Domain: trypsin homology <TRY>
F:31-161,49-65,133-234,140-207,172-186/Disulfide bonds: #status predicted
F:64,108,201/Active site: His, Asp, Ser #status predicted
F:76,78,81,86/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

```

alignment_scores:

```

Quality: 224.50 Length: 187
Ratio: 2.060 Gaps: 6
Percent Similarity: 58.289 Percent Identity: 30.481

```

alignment_block:

US-09-030-606-174 x A27547 ..

Align seg 1/1 to: A27547 from: 1 to: 247

```

3 TCAGCCGACACTGTTCCAGAGTGAGTCACAGGCTCTACACATCGG 52
||||| ::|||

```

```
61 SerAlaAlaHisCysTyrLysSerArgIleGln.....ValAr 73
53 GCTGGGCTGTCACAGTCTTGAGGCGCCAGACGAGGAGCCAGATGG 102
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
73 GluGlyGluHisAsnIleAspVal..ValGluGlyGlyGluGlnPheI 89
103 TGGAGGCGAGCCTTCCTACGCGCACCCAGAGTACACAGACCCCTGCTC 152
   ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
89 LeuAlaAlaLysIleIleArgHisProSerTyrAsnAlaAsnThrPhe 105
153 GCTACGACCTCATCTCATCAAGTTGACGATCCGTGCTCGAGCTGA 202
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
106 AspAsnAspIleMetLeuIleLysLeuAsnSerProAlaThrLeuAsnS 122
203 CACCATCCGAGACATCAGATGCTTCGCGAGTCCCTACCGGGGGAAC 252
   : : : : : ||||| ||||| ||||| ||||| ||||| |||||
122 ArgValSerThrValSerLeuProArgSerCysGlySerGlyThrL 139
253 CTGGCCTGCTTCTGCTGGGCTCTGCTGGCGAACGTGAGCTCAGGGT 302
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
139 ysCysLeuValSerGlyTyrGly..... 146
303 GTGTGCTGCCCCCTTCAAGAGAGTCTCTGCCAGTCGCGGGGCTGAC 352
   ::::: ||||| ||||| ||||| ||||| ||||| |||||
147 .....AsnThrLeuSerSer..GlyThrAsn 154
353 CCAGAGCTCTGGCTCCAGGCGAGATGCTACCGTCTGACATGGGTGA 402
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
155 .....TyrProSerLeuLeuGlnCysLeuAs 163
403 CGTGTGGGTGTGTCTGANGAGAGTCTGCAMTAAGCTTATGACCGCTGT 452
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
163 PAlaProValLeuSerAspSerSerCysLysSerSerTyrProGlyLysI 180
453 ACCACCCCANCAATGTTCTGCCCGCGGCGGAGGCGAAGACCGAAGACTCC 502
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
180 IeThrSerAsnMetPheCysLeuGlyPheLeuGlnGlyLysAspSer 196
503 TGCACAGCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 552
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
197 Cys.....GlnGlyAspSerGlyGlyProValValCysAsnGlyG 210
553 GAGAAGGGG 562
   ::|||
210 nleuGlnGly 213
```

This Page Blank (uspto)

OM of: US-09-030-606-174 to: SwissProt_37.* out_format : pfs
 Date: Sep 25, 1999 4:45 PM

About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=frimet-n2p.model -DEV=rlp
 -O=/cgr2_1/USPRO.spool/US09030606/runat_24091999_111618_29883/app_query.fasta.1
 -DB=SwissProt_37 -OPT=fastan -SUFFIX=isp -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -GAPOP=4.000 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=bloms62
 -TRANS=human40.cdt -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
 -ALIGN=15 -MODE=LOCAL -OUTEXT=pfs -NORM=stat -USER=US09030606
 -NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:

Query: us-09-030-606-174
 Query length: 1459
 Database: SwissProt_37.*
 Database sequences: 77977
 Database length: 28268293
 Search time (sec): 155.430000

Sequence	Strid	Orig	ZScore	Escore	Len	Documentation
SwissProt_37:ALU8_HUMAN	359.50	518.26	4.4e-22	591	P39195	homo sapiens (human)
SwissProt_37:ALU8_HUMAN	343.50	494.57	9.3e-21	581	P39195	homo sapiens (human)
SwissProt_37:ALU7_HUMAN	338.50	487.14	2.4e-20	593	P39194	homo sapiens (human)
SwissProt_37:ALU6_HUMAN	324.50	466.41	3.4e-19	593	P39193	homo sapiens (human)
SwissProt_37:ALU5_HUMAN	320.00	459.86	8.0e-19	585	P39192	homo sapiens (human)
SwissProt_37:ALU1_HUMAN	317.00	455.33	1.4e-18	591	P39188	homo sapiens (human)
SwissProt_37:ALU2_HUMAN	314.00	450.95	2.5e-18	587	P39189	homo sapiens (human)
SwissProt_37:ALU7_HUMAN	313.00	450.86	2.5e-18	593	P39194	homo sapiens (human)
SwissProt_37:ALU5_HUMAN	313.00	449.59	3.0e-18	585	P39192	homo sapiens (human)
SwissProt_37:ALU3_HUMAN	312.00	447.99	3.7e-18	587	P39190	homo sapiens (human)
SwissProt_37:ALU6_HUMAN	298.00	427.20	5.2e-17	591	P39188	homo sapiens (human)
SwissProt_37:ALU6_HUMAN	290.00	415.32	2.4e-16	593	P39193	homo sapiens (human)
SwissProt_37:ALU2_HUMAN	287.50	411.71	3.8e-16	587	P39189	homo sapiens (human)
SwissProt_37:ALU3_HUMAN	286.00	409.49	5.1e-16	587	P39190	homo sapiens (human)
SwissProt_37:ALU4_HUMAN	282.00	358.91	3.3e-13	603	P39191	homo sapiens (human)
SwissProt_37:TRV4_RAT	237.50	345.20	4.6e-12	246	P32821	rattus norvegicus (rat)
SwissProt_37:TRV3_RAT	237.50	345.20	4.6e-12	246	P32821	rattus norvegicus (rat)
SwissProt_37:TRV4_CANFA	236.00	342.98	6.2e-12	246	P06871	canis familiaris (dog)
SwissProt_37:TRV4_RAT	233.00	338.50	1.1e-11	247	P12788	rattus norvegicus (rat)
SwissProt_37:TRV1_BOVIN	226.50	329.02	3.7e-11	243	P00760	bos taurus (bovine)
SwissProt_37:TRV3_RAT	224.50	325.91	5.5e-11	247	P08426	rattus norvegicus (rat)
SwissProt_37:SCCE_HUMAN	222.00	322.01	8.8e-11	253	P49862	homo sapiens (human)
SwissProt_37:TRV2_XENLA	221.50	321.47	9.7e-11	244	P70059	xenopus laevis (atrid)
SwissProt_37:TRV1_HUMAN	221.50	321.47	9.7e-11	247	P70477	homo sapiens (human)
SwissProt_37:TRV3_SALSA	220.50	320.31	1.2e-10	238	P35033	salmo salar (atlantic)
SwissProt_37:TRV1_CHICK	220.50	319.96	1.2e-10	248	P08627	gallus gallus (chicken)
SwissProt_37:TRV2_CHICK	220.50	319.96	1.2e-10	248	P08628	gallus gallus (chicken)
SwissProt_37:TRV3_CHICK	219.00	317.74	1.6e-10	248	P08629	gallus gallus (chicken)
SwissProt_37:TRV2_SALSA	218.50	317.61	1.7e-10	231	P35032	salmo salar (atlantic)
SwissProt_37:TRV1_SALSA	213.50	309.80	4.4e-10	242	P35031	salmo salar (atlantic)
SwissProt_37:ALU4_HUMAN	214.00	302.64	4.4e-10	603	P39191	homo sapiens (human)
SwissProt_37:TRV1_XENLA	211.50	306.81	6.5e-10	243	P19799	xenopus laevis (atrid)
SwissProt_37:TRV4_HUMAN	211.50	304.87	6.6e-10	304	P35030	homo sapiens (human)
SwissProt_37:TRV1_HUMAN	208.50	302.26	1.1e-09	246	P00762	rattus norvegicus (rat)
SwissProt_37:TRV1_HUMAN	208.00	301.52	1.3e-09	246	P07166	mus musculus (mouse)
SwissProt_37:TRV1_PIG	207.50	301.32	1.4e-09	231	P00761	sus scrofa (pig)
SwissProt_37:TRV1_PIG	207.50	300.66	1.5e-09	229	P00764	squalus acanthias (sh)
SwissProt_37:TRV2_HUMAN	206.50	299.26	1.7e-09	247	P07478	homo sapiens (human)
SwissProt_37:TRV2_HUMAN	206.00	298.11	1.9e-09	259	P36376	rattus norvegicus (rat)
SwissProt_37:TRV2_RAT	202.00	292.63	3.9e-09	246	P00763	rattus norvegicus (rat)
SwissProt_37:KIKU_MOUSE	202.00	292.12	4.0e-09	261	P15946	mus musculus (mouse)
SwissProt_37:TRV2_BOVIN	197.00	285.19	1.0e-08	247	P29463	bos taurus (bovine)

seq_name: SwissProt_37:ALU8_HUMAN

seq_documentation_block:

ID ALU8_HUMAN STANDARD; PRT; 591 AA.

AC P39195.

DT 01-FEB-1995 (REL. 31, CREATED)

DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)

DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)

DE 1111 ALU SUBFAMILY SX WARNING ENTRY !!!!!

OS HOMO SAPIENS (HUMAN).

CC EURAROT4; METACOH; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

CC PRIMATES; CARNIVORIN; HOMINIDAE; HOMO.

CC (1)

CC SEQUENCE FROM N.A.

CC MEDLINE: 95021758.

CC CLAVIERE J.-M., KARALOWSKI W.;

CC "Alu alert.".

CC NATURE 371:752-752(1994).

CC (2)

CC CONCEPT.

CC MEDLINE: 92241891.

CC CLAVIERE J.-M.;

CC "Identifying coding exons by similarity search: alu-derived and other

CC potentially misleading protein sequences.";

CC GENOMICS 12:838-841(1992).

CC (3)

CC ALU FAMILIES CLASSIFICATION.

CC MEDLINE: 88333009.

CC OUBERTIN Y.;

CC "The Alu family developed through successive waves of fixation

CC closely connected with primate lineage history.";

CC J. MOL. EVOL. 27:194-202(1988).

CC (4)

CC ALU FAMILIES CLASSIFICATION.

CC MEDLINE: 91178815.

CC JUKKA J., MILOSAVLEVIC A.;

CC "Reconstruction and analysis of human Alu genes.";

CC J. MOL. EVOL. 32:105-121(1991).

CC -1- VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE THAT ALU REPEATS

CC FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS

CC SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES.

CC CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.

CC -1- ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON. 'xxx' IS USED

CC TO SEPARATE THE VARIOUS TRANSLATION PHASES.

CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER

CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO

CC ACID SEQUENCES.

CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND

CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE

CC ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN

CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,

CC cDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS

CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU

CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A

CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,

CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE

CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A

CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING

CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH

CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES

CC BEING REPORTED.

CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE

CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A

CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE

CC CODING NUCLEOTIDE SEQUENCE.

CC -----

CC This Swiss-Prot entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).

CC EMBL: U14574; -; NOT_ANNOTATED_CDS.

DR HYPOTHETICAL PROTEIN.
KW HYPOHETICAL PROTEIN.
FT DOMAIN 1 96 FRAME 1.
FT DOMAIN 2 100 195 FRAME 2.
FT DOMAIN 3 199 294 FRAME 3.
FT DOMAIN 4 298 393 FRAME 4.
FT DOMAIN 5 397 492 FRAME 5.
FT DOMAIN 6 496 591 FRAME 6.
SQ SEQUENCE 591 AA; 64395 MW; ABF69224 CRC32;

alignment_scores:
Quality: 359.50 Length: 96
Ratio: 4.609 Gaps: 1
Percent Similarity: 81.250 Percent Identity: 73.958

alignment_block:

US-09-030-606-174/rev x ALU8_HUMAN ..

Align seg 1/1 to: ALU8_HUMAN from: 1 to: 591

1447 TTTTGTGAGACAGCTTACTCTGTGCCCCAGGATGATGCTG 1398
446 PhepGluThiGluSerArgSerAlaGlnAlaGlyValGlnTrp 512
1397 TGATCTCACTCACTCACTCTGCTCCCATATTCAGCAATTCCT 1348
512 GAspGluGlySerLeuGlnProProProGlyPheGlyArgPheSer 529
1347 GCGTACGCTCCAGTACGCTGAGTACAGGCGCTGCCACATATCCA 1298
529 ySLeuSerLeuProSerSerTrpAspTrpArgAlaProProArgPro 545
1297 GCTACTTTGATTTTATTTAGTACAGAGGATTCACCATTTGGCCAGG 1248
546 AlaAsnPhcGlyIlePheSerArgSpGlyValSerProGlyTrpPro 562
1247 CTGGTCTTGAAGCTCTTACCTCAAGTGTGCTGCTGCTGCCCAAA 1198
562 yTrpSerArgTrpProAspLeuArg**Ser.....SerSerGlnS 576
1197 GTGCTGGAGTTACAGGACGACCGCTGCCAGCT 1160
576 eAlaGlyIleThrGlyValSerHisArgAlaArgPro 588

seq_name: SwissProt_37:ALU8_HUMAN

seq_documentation_block:

ID ALU8_HUMAN STANDARD; PRT; 591 AA.

AC P39195;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE !!! ALU SUBFAMILY SX WARNING ENTRY !!!
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95021758.
RA CLAVERIE J.-M., MAKALOWSKI W.:
RT "Alu alert."
RL NATURE 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE: 92241891.
RA CLAVERIE J.-M.:
RT "Identifying coding exons by similarity search: alu-derived and other

RT potentially misleading protein sequences.";

RL GENOMICS 12:838-841(1992).

RN [3]

RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE: 88333009.

RA OUELTIN Y.;

RT "The Alu family developed through successive waves of fixation

RT closely connected with primate lineage history.";

RL J. MOL. EVOL. 27:194-202(1998).

RN [4]

RP ALU FAMILIES CLASSIFICATION.

RX MEDLINE: 9117815.

RA JURKA J., MILOSAVLJEVIC A.;

RT "Reconstruction and analysis of human Alu genes.";

RL J. MOL. EVOL. 32:105-121(1991).

CC -1- VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS

CC FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS

CC SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES

CC CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.

CC -1- ISOLATED 'x' INDICATES THE PRESENCE OF A STOP CODON, 'xxx' IS USED

CC TO SEPARATE THE VARIOUS TRANSLATION PHASES.

CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER

CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO

CC ACID SEQUENCES.

CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND

CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE

CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN

CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,

CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS

CC LIGATED WITH ALU-DERIVED SEQUENCE IN AN ORIENTATION. ALTHOUGH ALU

CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A

CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,

CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE

CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A

CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING

CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH

CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES

CC BEING REPORTED.

CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE

CC WITH AN ALU TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A

CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE

CC CODING NUCLEOTIDE SEQUENCE.

CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).

CC EMBL: U14574; -; NOT_ANNOTATED_CDS.

DR HYPOTHETICAL PROTEIN.

KW HYPOHETICAL PROTEIN.

FT DOMAIN 1 96 FRAME 1.

FT DOMAIN 2 100 195 FRAME 2.

FT DOMAIN 3 199 294 FRAME 3.

FT DOMAIN 4 298 393 FRAME 4.

FT DOMAIN 5 397 492 FRAME 5.

FT DOMAIN 6 496 591 FRAME 6.

SQ SEQUENCE 591 AA; 64395 MW; ABF69224 CRC32;

alignment_scores:

Quality: 343.50 Length: 97

Ratio: 4.294 Gaps: 1

Percent Similarity: 82.474 Percent Identity: 72.165

alignment_block:

US-09-030-606-174 x ALU8_HUMAN ..

Align seg 1/1 to: ALU8_HUMAN from: 1 to: 591

1157 AAGAGGCTGGCAGGAGGCTGCTGCTGATCCAGCACTTTGGAGG 1206

```

      3 ArgArgProGlyAlaValAlaHisAlaCysAsnProSerThrIleuLys.. 18
1207 CGAGCGAGCGAGTACTGAGTGAAGAGTTGAGCCAGCGCTGGCCAA 1256
      19 .....GlyArgSerProGlyValAlaArgSerSerArgProAlaTrrPro 33
1257 AATGGTGAATACCTGCTGTACTATAAAATAACAAGTTAGCTGAGATATG 1306
      33 htrTp**AsnProValSerThrLysAsnThrLysIleSerArgAlaTrr 49
1307 TGGCAGGCGCCCTGTATCCAGCTACTTGGGAGCGCTGAGCGAGAAAT 1356
      50 TTPArgAlaProValIleProAlaTrrArgGluAlaGluAlaGlyGlu 66
1357 GCTGAATATGGAGCGAGGTGAAGTGAAGTGAATCAGCCACTAT 1406
      66 rIeuGluProGlyArgArgArgLeuGln**AlaGluIleAlaProLeuH 83
1407 ACTCCAGCTGGGCGCAAGAGATGAAGTCTGTCTCAAAAAA 1447
      83 lSerSerIeuGlyAspArgAlaArgLeuArgLeuLysLys 96

seq_name: SwissProt_37:ALU7_HUMAN

seq_documentation_block:
ID ALU7_HUMAN STANDARD; PRT; 593 AA.
AC P39194;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DE 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE 1111 ALU SUBFAMILY SO WARNING ENTRY 1111
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95021758.
RA CLAVERIE J.-M., MAKALOWSKI W.;
RT "Alu alert.";
RL NATURE 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE: 92241891.
RA CLAVERIE J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL GENOMICS 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE: 88333009.
RA QUENTIN Y.;
RT "The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
RL J. MOL. EVOL. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE: 91178815.
RA JURKA J., MILOSAVLJEVIC A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. MOL. EVOL. 32:105-121(1991).
CC -1- VARIOUS ANALYZES (SEE REF. 3 AND REF. 4) INDICATE THAT ALU REPEATS
CC FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS
CC SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES
CC CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
CC -1- ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'XXX' IS USED
CC TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN

```

```

CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14573; -; NOT_ANNOTATED_CDS.
KW HYPOTHETICAL PROTEIN.
FT DOMAIN 1 97 FRAME 1.
FT DOMAIN 2 101 196 FRAME 2.
FT DOMAIN 3 200 295 FRAME 3.
FT DOMAIN 4 299 395 FRAME 4.
FT DOMAIN 5 399 494 FRAME 5.
FT DOMAIN 6 498 593 FRAME 6.
SQ SEQUENCE 593 AA; 64417 MW; 56C24239 CRC32;

alignment_scores:
      Quality: 338.50      Length: 97
      Ratio: 4.285
Percent Similarity: 81.443 Percent Identity: 71.134

alignment_block:
US-09-030-606-174/rev x ALU7_HUMAN
Align seg 1/1 to: ALU7_HUMAN from: 1 to: 593

1447 TTTTGTGACAGACAGTCTACTCT...GTGGCCAGCTGAGTATAGTG 1401
      ||||| ..... ||||| ||||| ||
299 PhePheLeuArgArgSerPheAlaLeuValAlaGlnAlaGlyValGlnTr 315
1400 GTGTGATCTCAACCTCAACGCTGCGCTGCCATATTCAGCAATCT 1351
      | |||||..... ||||| |||||..... |||||
315 PArgAspLeuGlySerLeuGlnProProProGlyPheLysArgPheS 332
1350 CCTGCTCAGCCCTCCAGTAGCTGGATTACAGCGCGCTGCCACCATAT 1301
      ||||| ||||| ||||| ||||| ||||| |||||
332 ercYLeuSerLeuProSerSerTrpAspTrpArgArgProProArg 348
1300 CCAGCTACTTTGTATTATTATAGTACAGACAGGATTTCACATTTGGCC 1251
      ||||| ||||| ||||| ||||| ||||| |||||
349 ProAlaAsnPhcCysIlePheSerArgAspGlyValSerProCysTrp 365
1250 AGCGTGGTCTGAACCTTACCTTACCAAGAGATGCGCGCCGCG. CTC 1202
      ||||| ||||| ||||| ||||| ||||| |||||
365 oGlyTrpSerArgTrpProAspLeuArg**SerThrArgLeuGlyLeuP 382
1201 CAAAGTGTGGATTACAGAGCATGAGCCAGCCCTGCCAGCC 1161
      ||||| ||||| ||||| ||||| ||||| |||||
382 rOlyCysTrpAspTrpArgArgGluProProArgProAla 395

seq_name: SwissProt_37:ALU6_HUMAN

seq_documentation_block:
ID ALU6_HUMAN STANDARD; PRT; 593 AA.

```



```

RL J. MOL. EVOL. 27:194-202(1988).
RN [4]
RX ALU FAMILIES CLASSIFICATION.
RA MEDLINE: 91178815.
RT JURKA J., MILOSAVLJEVIC A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. MOL. EVOL. 32:105-121(1991).
CC -1- VARIOUS ANALYZES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS
CC FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS
CC SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES
CC CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
CC -1- ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'xxx' IS USED
CC TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGHTED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U14571; ; NOT_ANNOTATED_CDS.
CC DR KW HYPOTHETICAL PROTEIN
CC FT DOMAIN 1 95 FRAME 1.
CC FT DOMAIN 2 99 FRAME 2.
CC FT DOMAIN 3 197 FRAME 3.
CC FT DOMAIN 4 295 FRAME 4.
CC FT DOMAIN 5 393 FRAME 5.
CC FT DOMAIN 6 491 FRAME 6.
CC FT DOMAIN 7 585 FRAME 7.
CC SQ SEQUENCE 585 AA: 63957 MW: 5C794090 CRC32:

alignment_scores:
    Quality: 320.00 Length: 97
    Ratio: 4.000 Gaps: 2
Percent Similarity: 82.474 Percent Identity: 71.134

alignment_block:
US-09-030-606-174/rev x ALU5_HUMAN ..
Align seg 1/1 to: ALU5_HUMAN from: 1 to: 585

1447 TTTTGTGAGACAGCTTACTCTGTTGCCCCAGCTGAGTATAGTGTG 1398
|||||
491 PheheglutthrglucSerValala.ArgleuclucysSergerly 507
1397 TGAGCTTAACCACTTCACTGCGCTCCCATATTCAGAAATTCCT 1348
|||||
507 lalleseAlahiscysasnleuArgleuProglySerSerAspsePro 523
1347 GCCTACGCTCCCAAGTAGTGTGGATTACAGCGCCCTGCACCATATCA 1298

```

```

|||||
524 AlaseralaserArgValalaGlyThrThrglyAlaArgIshislaG 540
1297 GCTACTTTTGATTTTATTACAGACAGATTTTCACCATTTGGCCAGG 1248
|||||
540 nleuilePheValaPheleuValaGluThrIrglyPhehisIshisValaGlylna 557
1247 CTGCTGTGAACCTTACTACCTCAAGTAGTGTGCTGCGCTGCCCAA 1198
|||||
557 spglyleuAspIleuThrSer***SerAlaArgleuGly.LeuProly 573
1197 GTGCTGGATTACAGCATGAGCCACCCTGCCAGCC 1161
|||||
573 scystirAspIlyArgIrglyuProArgProAla 585

seq_name: SwissProt_37:ALU1_HUMAN
seq_documentation_block:
ID ALU1_HUMAN STANDARD; PRT; 591 AA.
AC P39188;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE |||| ALU SUBFAMILY J WARNING ENTRY ||||
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA.
RN [1]
RP PRIMATE; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95021758.
RA CLAVIERIE J.-M., MAKALOWSKI W.;
RT "Alu alert.";
RL NATURE 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE: 92241891.
RA CLAVIERIE J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL GENOMICS 12:838-841(1992).
RN [3]
RX ALU FAMILIES CLASSIFICATION.
RP MEDLINE: 88333009.
RA QUENTIN Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. MOL. EVOL. 27:194-202(1988).
RN [4]
RX ALU FAMILIES CLASSIFICATION.
RA MEDLINE: 91178815.
RT JURKA J., MILOSAVLJEVIC A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. MOL. EVOL. 32:105-121(1991).
CC -1- VARIOUS ANALYZES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS
CC FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS
CC SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES
CC CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
CC -1- ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'xxx' IS USED
CC TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGHTED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING

```

DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES BEING REPORTED.

-1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A PART OF ALU REPEAT MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE CODING NUCLEOTIDE SEQUENCE.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL, U14567; -, NOT_ANNOTATED_CDS.

DR HYPOTHEITICAL PROTEIN.

FT	DOMAIN	1	96	FRAME 1.
FT <td>DOMAIN</td> <td>100</td> <td>195</td> <td>FRAME 2.</td>	DOMAIN	100	195	FRAME 2.
FT <td>DOMAIN</td> <td>199</td> <td>294</td> <td>FRAME 3.</td>	DOMAIN	199	294	FRAME 3.
FT <td>DOMAIN</td> <td>298</td> <td>393</td> <td>FRAME 4.</td>	DOMAIN	298	393	FRAME 4.
FT <td>DOMAIN</td> <td>397</td> <td>492</td> <td>FRAME 5.</td>	DOMAIN	397	492	FRAME 5.
FT <td>DOMAIN</td> <td>496</td> <td>591</td> <td>FRAME 6.</td>	DOMAIN	496	591	FRAME 6.
SO <td>SEQUENCE</td> <td>591 AA;</td> <td>63790 MW;</td> <td>AA089112 CRC32;</td>	SEQUENCE	591 AA;	63790 MW;	AA089112 CRC32;

alignment_scores:

Quality:	317.00	Length:	96
Ratio:	4.284	Gaps:	0
Percent Similarity:	77.083	Percent Identity:	69.792

alignment_block:

US-09-030-606-174/rev x ALU1_HUMAN ..

Align seg 1/1 to: ALU1_HUMAN from: 1 to: 591

```

1447 TTTTGTGAGACAGAGCTCTGCTGTCGCCAGCTGAGATATAGTGTG 1398
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
496 PhePhgdluThrclYserArgseValAlaGlnAlaGlyValAlntrpar 512
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1397 TGATGTCACTACTTCACTTCACTTGCCTCCCATATTCAGCAATTCCT 1348
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
512 GAAPHISGLYserLeuGlnProArGProGlyLeuLysArgSerSerc 529
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1347 GCCTGAGCCGCCAAGTAGCTGGATTACAGGGCCGCCACCATATCCA 1298
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
529 YsluSerLeuProSerSerTrpAspTrpArgAlaProProArGPro 545
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1297 GCTAAGTTTGTATTTTATGACAGACAGATTTCACATTGGCCAGG 1248
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
546 AlaAspPhcysIlePhcysArgAspGlyValSerLeuLysProG 562
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1247 CTGGCTTGAATCTCTTACCTCAAGTGATGCTGCTGCTGCG. CTCCCA 1199
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
562 YTPSerArgThrProGlyLeuLys**SerSerArgLeuGlyLeuProL 579
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1198 AGTGTGAGGATTACAGGACGACCCGCCAGCC 1161
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
579 YscYstrpAspTrpArgArgGlnProProArGProAla 591
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

seq_name: SwissProt_37:ALU2_HUMAN

seq_documentation_block:

ID	ALU2_HUMAN	STANDARD:	PRT:	587 AA.
AC	P39189;			
DT	01-FEB-1995 (REL. 31, CREATED)			
DT	01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)			
DT	01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)			
DE	1111 ALU SUBFAMILY SB WARNING ENTRY 1111			
OS	HOMO SAPIENS (HUMAN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;			
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.			

[1] SEQUENCE FROM N.A.

RA MEDLINE; 95021758.

RA CLAVERIE J.-M., MAKALOWSKI W.;

RT "Alu alert.";

RL NATURE 371:752-752(1994).

[2] CONCEPT.

RA MEDLINE; 92241891.

RA CLAVERIE J.-M.;

RT "Identifying coding exons by similarity search: alu-derived and other potentially misleading protein sequences.";

RL GENOMICS 12:838-841(1992).

[3] ALU FAMILIES CLASSIFICATION.

RP MEDLINE; 86333009.

RA QUENTIN Y.;

RT "The Alu family developed through successive waves of fixation closely connected with primate lineage history.";

RL J. MOL. EVOL. 27:194-202(1988).

[4] ALU FAMILIES CLASSIFICATION.

RP MEDLINE; 9117815.

RA JURKA J., MILOSAVLJEVIC A.;

RT "Reconstruction and analysis of human Alu genes.";

RL J. MOL. EVOL. 32:105-121(1991).

-1- VARIOUS ANALYZES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.

-1- ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.

-1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO ACID SEQUENCES.

-1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER, CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS, CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES BEING REPORTED.

-1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A PART OF ALU REPEAT MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE CODING NUCLEOTIDE SEQUENCE.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL, U14568; -, NOT_ANNOTATED_CDS.

DR HYPOTHEITICAL PROTEIN.

FT	DOMAIN	1	96	FRAME 1.
FT <td>DOMAIN</td> <td>100</td> <td>194</td> <td>FRAME 2.</td>	DOMAIN	100	194	FRAME 2.
FT <td>DOMAIN</td> <td>198</td> <td>292</td> <td>FRAME 3.</td>	DOMAIN	198	292	FRAME 3.
FT <td>DOMAIN</td> <td>296</td> <td>391</td> <td>FRAME 4.</td>	DOMAIN	296	391	FRAME 4.
FT <td>DOMAIN</td> <td>395</td> <td>489</td> <td>FRAME 5.</td>	DOMAIN	395	489	FRAME 5.
FT <td>DOMAIN</td> <td>493</td> <td>587</td> <td>FRAME 6.</td>	DOMAIN	493	587	FRAME 6.
SO <td>SEQUENCE</td> <td>587 AA;</td> <td>63703 MW;</td> <td>844F0DCE CRC32;</td>	SEQUENCE	587 AA;	63703 MW;	844F0DCE CRC32;

alignment_scores: length: 95
 quality: 314.00
 ratio: 4.486
Percent Similarity: 73.684 Percent Identity: 70.526

alignment_block:
US-09-030-606-174/rev x ALU2_HUMAN

Align seg 1/1 to: ALU2_HUMAN from: 1 to: 587

```
1447 TTTTGGAGAGAGCTTACTCTGTTCCAGCTGAGATAGTGTG 1398
|||||
493 Phepegutrhgluserararaladgnalaglyvalgntprpar 509
1397 TGATCTCACTTCACTGCTGCTCCATATTCAGCAATCTCT 1348
|||||
509 gdspleuclyserleuglnalaproprologlypherhpropheserc 526
1347 GCCTCAGCTCCCAAGTAGTGGATTACAGCGCGCTGCCACCATATCCA 1298
|||||
526 yslaserleuprosersertirpapyrargrproprobrpro 542
1297 GCTACTTTTATTTTATTTAGTACAGAGATTGACATTTGGCCAGG 1248
|||||
543 Alasnphecysilepheserargaspelyalserprocy**Progl 559
1247 CTGCTCTGAGACTCTTACTCAAGTGTATCTGCTGCTGCCCTCCCAA 1198
|||||
559 ytrserargserproaspleu.valleargproprobrpro 575
1197 GTGCTGGATTACAGCATGAGCCACCTGCC 1165
|||||
576 Valleuglyleuglnala**Alathralapro 586
```

seq_name: SwissProt_37:ALU7_HUMAN

```
seq_documentation_block:
ID ALU7_HUMAN STANDARD; PRT; 593 AA.
AC P39194;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DE 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE 1111 ALU SUBFAMILY SQ WARNING ENTRY 1111
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE: 95021758.
RA CLAVERIE J.-M., MAKALOWSKI W.;
RT "Alu alert.";
RL NATURE 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE: 92241891.
RA CLAVERIE J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL GENOMICS 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE: 88333009.
RA QUENTIN Y.;
RT "The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
RL J. MOL. EVOL. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE: 91178815.
RA JURKA J., MILOSAVLJEVIC A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. MOL. EVOL. 32:105-121(1991).
CC -1- VARIOUS ANALYZES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS
```

```
CC FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS
CC SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES
CC CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
CC ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'XXX' IS USED
CC TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -1- CAUTION: ALU REPERITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTRAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU REPEAT ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U14573; -; NOT_ANNOTATED_CDS.
CC KW HYPOTHETICAL PROTEIN.
CC FT DOMAIN 1 97 FRAME 1.
CC FT DOMAIN 101 196 FRAME 2.
CC FT DOMAIN 200 285 FRAME 3.
CC FT DOMAIN 299 385 FRAME 4.
CC FT DOMAIN 399 494 FRAME 5.
CC FT DOMAIN 498 593 FRAME 6.
CC SO SEQUENCE 593 AA; 64417 MW; 56C24239 CRC32;
```

alignment_scores: length: 97
 quality: 314.00
 ratio: 4.026
Percent Similarity: 80.412 Percent Identity: 70.103

alignment_block:
US-09-030-606-174 x ALU7_HUMAN

Align seg 1/1 to: ALU7_HUMAN from: 1 to: 593

```
1161 GCGTGGCAGGCTGCTCATGCCCTGATCCAGCATTTGGGAGGC GA 1209
|||||
1 GlyceralalargtrpleurnProvalilleProalaleutrglnaagl 17
1210 GCGAGCAGATCACTTGAGTGAAGAGTTCAGACCGCGGCAAAAT 1259
|||||
17 ualaglyglserprogluvalargserargprovalatrtprothrt 34
1260 GGTGAAATCTGTCTGTACTAAATAACAAAGTTAGCTGATGTGCG 1309
|||||
34 rp**AsnprovalserThrlysnthrlyslleserargalatriptp 50
1310 CAGCGCCTGTATCCAGCATCTGGAGCGTGGAGCAGCAGCATGTCT 1359
|||||
51 ArgalaprovalilleProvalatThrarggluvalaglyglserle 67
1360 TGATATGGAGGAGCAGAGTGAAGTAGTGAAGTACACACACTATACT 1409
```



```
|||||:::|||||
265 rogiylglylacysserguProargSerArgHiscysthrProAla 281
1416 GGGGCAACAGACTGTCTCTCAAAAA 1448
282 TrpAlaThrGluArgAspSerAlaSerIlyS 292

seq_name: SwissProt_37:ALU4_HUMAN

seq_documentation_block:
ID ALU4_HUMAN STANDARD; PRT; 603 AA.
AC P39191;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE ||||| ALU SUBFAMILY SB2 WARNING ENTRY |||||
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95021758.
RA CLAVERIE J.-M., MAKALOWSKI W.;
RT "Alu alert.";
RL NATURE 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE; 92241891.
RA CLAVERIE J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL GENOMICS 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE; 88333009.
RA QUENENIN Y.;
RT "The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
RL J. MOL. EVOL. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE; 91178815.
RA JURKA J., MILOSAVLJEVIC A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. MOL. EVOL. 32:105-121(1991).
-1- VARIOUS ANALYZES (SEE REF.3 AND REF.4) INDICATE THAT ALU REPEATS
FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS
SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES
CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.
-1- ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'XXX' IS USED
TO SEPARATE THE VARIOUS TRANSLATION PHASES.
-1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
ACID SEQUENCES.
-1- CAUTION: ALU REPERITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN
ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
BEING REPORTED.
-1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
PART OF ALU REPEAT MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE
CODING NUCLEOTIDE SEQUENCE.
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U14570; -; NOT_ANNOTATED_CDS.
KW HYPOTHETICAL PROTEIN.
FT DOMAIN 1 98 FRAME 1.
FT DOMAIN 102 199 FRAME 2.
FT DOMAIN 202 300 FRAME 3.
FT DOMAIN 304 401 FRAME 4.
FT DOMAIN 405 502 FRAME 5.
FT DOMAIN 506 603 FRAME 6.
SQ SEQUENCE 603 AA; 65272 MW; A6C07F68 CRC32;
```

```
alignment_scores:
Quality: 252.00 Length: 99
Ratio: 3.150 Gaps: 2
Percent Similarity: 80.808 Percent Identity: 63.636
```

```
alignment_block:
US-09-030-606-174/rev x ALU4_HUMAN ..
```

```
Align seq 1/1 to: ALU4_HUMAN from: 1 to: 603
```

```
1447 TTTTGGAGACAGAGTCTTACTGTGTCGCCAGCTGGAGTATA..... 1404
|||||
506 PheheglutrhrgluserArgSerValAlaGlnAlaGlyLeuAlaGlyThrAl 522
1403 .GGGTGTGATCTCAACTCACTTCAACCTGTGCTCCCAATTTCAAGCAA 1355
|||||
522 aValAlaGlnSerArgLeuThrAlaSerSerAlaSerArgValAlaAlaI 539
1354 TTCCTTCGCCCTCAGCCCTCCCAAGTGGATTCAGGCGCCCTGCCACC 1305
|||||
539 leLeuLeuProGlnProProGln**LeuGlyLeuGlnAlaProAlaThr 555
1304 ATATCAGACTA.ACTTTGTATTTTGTAGTACAGACAGATTCACCAATT 1256
|||
556 AlaProGly***PhePheValPheLeuValGluThrGlyPheHisLeuVal 572
1255 TGCCAGAGCTGCTTGAACCTCTTACCTCAAGTGAATCGCTGCTCGC 1206
|||
572 lSerGlnAspGlyLeuAspLeuThrSer***SerThrArgLeuGly. 588
1205 CTCCTCAAGTCTGGATTCAGGCAATGAGCCACCTGCCAGCC 1161
|||||
589 leuProLysCysTirPaspIyArgArgGlnProProAla 603
```

OM of: US-09-030-606-174 to: SPTRMBL_10:* out_format: pfs
Date: Sep 25, 1999 10:35 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:

-MODEL=facet+np.model -DEV=xlp
-O=/gen2.1/USPO_spo0/US09030606/runat_24091999_171617_29669/app-query.fasta.1
-DB=SPTRMBL_10 -QPR=fastcan -SUFFIX=rspt -GAPO=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000
-GAPO=6.500 -GAPEXT=0.050 -XGAPO=10.000 -XGAPEXT=0.500
-GAPO=6.000 -GAPEXT=7.000 -XGAPO=10.000 -XGAPEXT=0.500
-DELDP=6.000 -DELXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdt -LIST=45 -DOCALLIN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:

Query: US-09-030-606-174
Query length: 1459
Database: SPTRMBL_10:*
Database sequences: 201082
Database length: 61543640
Search time (sec): 260.540000

```
score_list:
Sequence      Strd Orig      ZScore      EScore      Len      Documentation
SP_rident:Q920M1 + 477.50      835.71      4.7e-39      255      1 Q920M1 mus musculus (mouse). en
SP_human:Q78525 - 310.00      529.12      1.2e-22      666      1 P78525 homo sapiens (human). my
SP_human:Q14608 - 280.00      471.61      2.1e-19      1079      1 Q14608 homo sapiens (human). u
SP_human:Q99770 - 270.50      471.80      1.6e-18      139      1 Q99770 homo sapiens (human). hy
SP_human:Q60448 - 268.00      459.05      3.0e-18      375      1 Q60448 homo sapiens (human). ne
SP_human:Q15662 - 257.00      439.60      3.7e-17      368      1 Q15662 homo sapiens (human). tr
SP_invertebrate:Q26195 - 247.00      426.87      3.5e-16      200      1 Q26195 Plasmodium vivax. pval
SP_human:Q07826 - 246.50      431.77      3.7e-16      100      1 Q07826 homo sapiens (human). x
SP_hmc:Q29976 - 242.00      418.13      1.1e-15      196      1 Q29976 homo sapiens (human). ma
SP_invertebrate:Q92099 + 221.00      378.93      1.4e-13      142      1 Q92099 paratubercula magellani
SP_invertebrate:Q93265 + 220.00      376.88      1.7e-13      250      1 Q93265 pleuroctes americanus
SP_invertebrate:Q91515 + 219.50      376.13      1.9e-13      237      1 Q91515 fuqu rubripes (japanese
SP_invertebrate:Q92046 + 219.00      375.13      2.1e-13      249      1 Q92046 disostichus mawsoni. pr
SP_invertebrate:Q93266 + 213.50      365.56      7.5e-13      242      1 Q93266 pleuroctes americanus
SP_invertebrate:Q93594 + 212.50      366.35      9.3e-13      178      1 Q93594 dienerichus labrax (eu
SP_human:Q15665 - 211.00      361.43      1.2e-12      259      1 Q15665 homo sapiens (human). tr
SP_rident:Q63274 + 211.00      361.35      1.3e-12      235      1 Q63274 rattus norvegicus (rat).
SP_human:Q75837 + 210.00      359.83      1.4e-12      282      1 Q75837 rattus norvegicus (rat).
SP_rident:Q92189 + 210.00      359.19      1.7e-12      246      1 Q92189 mus musculus (mouse). tr
SP_human:Q15946 - 207.00      345.75      1.0e-11      222      1 Q15946 homo sapiens (human). gl
SP_invertebrate:Q42160 + 201.50      344.07      1.2e-11      245      1 Q42160 petromyzon marinus (sea
SP_invertebrate:Q42608 + 201.50      344.00      1.2e-11      247      1 Q42608 petromyzon marinus (sea
SP_rident:Q63275 + 200.50      342.50      1.5e-11      239      1 Q63275 rattus norvegicus (rat).
SP_invertebrate:Q42158 + 200.50      342.22      1.5e-11      250      1 Q42158 petromyzon marinus (sea
SP_rident:Q03955 + 199.50      340.34      1.7e-11      250      1 Q03955 prionyx natlensis (afri
SP_invertebrate:Q41036 + 196.50      336.10      3.7e-11      219      1 Q41036 gadus morhua (atlantic d
SP_invertebrate:Q42159 + 193.50      333.41      4.6e-11      244      1 Q42159 petromyzon marinus (sea
SP_mammal:Q29474 - 191.00      324.83      1.3e-10      261      1 Q29474 canis familiaris (dog).
SP_human:Q13675 - 189.50      316.74      1.9e-10      499      1 Q13675 homo sapiens (human). al
SP_human:Q15096 + 187.50      319.50      2.3e-10      234      1 Q15096 homo sapiens (human).
SP_invertebrate:Q08309 + 186.50      316.81      3.6e-10      261      1 Q08309 mus musculus (mouse). pr
SP_invertebrate:Q18599 + 186.00      316.08      4.1e-10      256      1 Q18599 drosophila villis (fr
SP_rident:Q61855 + 183.50      315.02      4.6e-10      261      1 Q61855 mus musculus (mouse). fr
SP_human:Q13629 + 184.50      319.87      5.4e-10      118      1 Q13629 homo sapiens (human). hy
SP_rident:Q05484 + 173.50      293.96      7.1e-09      251      1 Q05484 rattus norvegicus (rat).
SP_rident:Q61388 + 172.50      292.14      8.9e-09      252      1 Q61388 mus musculus (mouse). cy
SP_invertebrate:Q27761 + 172.50      291.69      8.9e-09      266      1 Q27761 penaeus vannamei (pend
SP_invertebrate:Q07620 + 170.50      288.51      1.4e-08      264      1 Q07620 stomoxys calcitrans (s
SP_invertebrate:Q62361 + 169.50      286.41      1.8e-08      264      1 Q62361 penaeus vannamei (pend
SP_invertebrate:Q01309 + 169.00      286.21      2.0e-08      243      1 Q01309 botryllus schlosseri.
SP_human:Q95809 - 168.00      296.69      2.2e-08      56      1 Q95809 homo sapiens (human). de
SP_rident:Q08301 + 165.50      279.87      4.4e-08      246      1 Q08301 mus musculus (mouse). se
```

```
SP_invertebrate:Q91053 + 162.00      273.10      9.8e-08      262      1 Q91053 glycydinus ussuriensis.
SP_mammal:Q46683 + 160.50      270.79      1.4e-07      251      1 Q46683 ovis aries (sheep). m
SP_invertebrate:Q9Y6S1 + 159.50      269.59      1.7e-07      234      1 Q9Y6S1 agkistrodon acutus (d
seq_name: SP_rident:Q920M1
```

seq_documentation_block:

```
ID      Q920M1      PRELIMINARY:      PRT:      255 AA.
AC      Q920M1:
DT      01-MAY-1999 (TREMBLrel. 10, Created)
DT      01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT      01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE      ENAMEL MATRIX SERINE PROTEINASE 1 PRECURSOR.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC      Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=SWISS-WEBSER;
RA      Slammer J.;
RT      *Enamel Matrix Serine Proteinase 1 (EMSP1).";
RL      Submitted (Aug-1997) to the EMBL/Genbank/DBD databases.
DR      EMBL; AF019979; AAC98894.1; -.
KW      Signal.
FT      SIGNAL.
FT      CHAIN.
FT      MATRIX SERINE PROTEINASE 1.
SQ      SEQUENCE      255 AA.      27488 MW;      6E711616 CRC32;
```

alignment_scores:

```
Quality: 477.50      Length: 173
Ratio: 3.820      Gaps: 3
Percent Similarity: 72.254      Percent Identity: 55.491
```

alignment_block:

```
US-09-030-606-174 x Q920M1 ..
```

```
Align seg 1/1 to: Q920M1 from: 1 to: 255
```

```
3 TCAGCCGACACCTGTTCCAGAGTGCAGAGCTCTACACCATCGG 52
|||||
69 SerAlaIahScys.....LeuGlnGlnSerTyrIleValGI 81
53 GCTGGGCGCTCAGAGCTGTAGCGCCAGCAAGCCAGGAGCAAGTGG 102
|||||
81 YleGlnGlnHisHisLeuYsgIserGlnGlnProGlySerTargMetL 98
103 TGGAGCGCAGCTCTCCGTAAGCGCAGCAGATACACACCTTCTC 152
|||||
98 eugIAlahHisLeuSerIleGlnHisProAsnPhaAsnAProSerPhe 114
153 GCTACAGACCTCATGCTCATCAAGTTGGAGCAATCGTGTCCAGTCTGA 202
|||||
115 AlasnaSpLeuMetIleuIleYsLeuAsnGlnSerValIleGlnSerAs 131
203 CAGCATCCGAGACATGACATGCTGTCGACGTCCTACCGCGGAGACT 252
|||||
131 nThIleYsSerIleProValAlaTrnGlnCysProTrnProGlyIAspt 148
253 CTTGCTCGTGTTCGCTGCGGCTCTGCGCAACGCTGACGCTCAGCGGT 302
|||||
148 hrcYsLeuValSerGlyTyrProGlnLeuYsAsnGly..Lys..... 161
303 GTGTGTCTGCCCTTTCAGAGAGGTCTTCCAGTCTGCGGGGGGTGAC 352
|||||
161 .....
353 CCAGAGCTCTGCGCTCCAGGACGAATGCTGCTGAGTGCCTGAGAA 402
|||||
162 .....
403 CGTGTGCTGTGCTGCTGANGAGGCTGCTGACCTTAAGCTTATGACCGCTGT 452
|||||
```

170 mLeuSerValAlaSerGluGluThrCysArgLeuLeuTyrAspProVal 187
 453 ACCAGCCCANCATGTTGCGCGGAGGAGGAGACGACGAGAGACTCC 502
 187 yRh1sLeuSerMetPheCysAlaGlyGlyGlyAlaSpSer 203
 503 TGCACGCTGAGAGAGGG 520
 204 CysAnGlyAspSerGly 209
 seq_name: sp_human:P78525

seq_documentation_block:
 ID P78525 PRELIMINARY; PRT; 666 AA.
 AC P78525; P78526;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE MYB PROTO-ONCOGENE PROTEIN (C-MYB).
 GN C-MYB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER, AND PLACENTA;
 RA WESTIN E.H., GORSE K.M.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U22376; AAB49034.1; -.
 DR EMBL; U22376; AAB49037.1; 3.
 DR PROSITE; PS00037; MYB_1; 3.
 DR PROSITE; PS00334; MYB_2; 3.
 DR PFM; PF00249; myb_DNA-binding; 3.
 KW Proto-oncogene; Nuclear protein; DNA-binding; Repeat;
 KM Transcription regulation; Activator.
 FT DNA_BIND 34 86 MYB.
 FT DNA_BIND 87 138 MYB.
 FT DNA_BIND 139 189 MYB.
 SQ SEQUENCE 666 AA; 75525 MW; 2E666087 CRC32;

alignment_scores:
 Quality: 310.00 Length: 97
 Ratio: 3.827 Gaps: 1
 Percent Similarity: 83.505 Percent Identity: 71.134

alignment_block:
 US-09-030-606-174/rev x P78525 ..

Align seg 1/1 to: P78525 from: 1 to: 666

1447 TTTTGGACAGAGCTTACTGTGCCCCAGAGTAGATAGTGTG 1398
 556 PheThGlnThrSerProValAlaAspAlaProThrGlyValGlnTrpH 572
 1397 TGATCTCAACTCACTTCAACCTGCTCCCATATTCAGCAATTCCT 1348
 572 sAspPheGlySerLeuGlnProLeuProGlyPheLeuArgPheSer 589
 1347 GCTCAGCCCTCCCAAGTAGTGGGATTACAGCGCCTGCCACCATTC 1298
 589 yLeuSerLeuProArgSerTrpAspTyrGlnHisProProArgPro 605
 1297 GCTAATTTTGTATTTAGTACAGAGAGATTTCACCATTTGGGCGAG 1248
 606 AlaAsn.PheGlnPheLeuValGlnThrGlyPheLeuHisValGlyGln 622
 1247 CTGGCTTGAAGCTTACTTCAAGTAGATGCTGCTGCTC. GCTGCCAA 1199
 622 laGlyLeuGlnLeuLeuThrSerGlyAspLeuProAlaSerAlaSerGln 638
 1198 AGTGTGGGATTACAGGATGAGCCACCTGCCAGGCT 1160

639 SerAlaArgIleThrGlyValSerHisArgAlaArgPro 651

seq_name: sp_human:O14608
 seq_documentation_block:
 ID O14608 PRELIMINARY; PRT; 1079 AA.
 AC O14608;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE UBQUITOUS TPR MOTIF, Y ISOFORM.
 GN UTY.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 98022381.
 RA LAHN B.T., PAGE D.C.;
 RT "Functional coherence of the human Y chromosome.";
 RL Science 278:675-680(1997).
 DR EMBL; AF000986; AAC51843.1; -.
 DR PFM; PF00515; TPR; 5.
 SQ SEQUENCE 1079 AA; 118073 MW; D7A4F281 CRC32;

alignment_scores:
 Quality: 280.00 Length: 85
 Ratio: 3.836 Gaps: 1
 Percent Similarity: 85.882 Percent Identity: 72.941

alignment_block:
 US-09-030-606-174/rev x O14608 ..

Align seg 1/1 to: O14608 from: 1 to: 1079

1414 GCTGAGTATAGTGGTGTGATCTCAACTCACTTCAACCTGCTCCCAT 1365
 996 AlaGlyMetGlnTrpCysAspLeuSerLeuGlnProProProG 1012
 1364 ATTCAAGCAATTCCTCTGCTCAAGCTCCCAAGTAGCGGATTACAGGC 1315
 1012 yPheLeuArgPheSerHisLeuSerLeuProAsnSerTrpAsnTyrArgH 1029
 1314 GCTGCCACCATATCCAGCTTACTTTGTATTTAGTACAGAGGATT 1265
 1029 isLeuProSerCysProThrAsnPheCysIlePhe.ValGlnThrGlyPh 1045
 1264 TCACATTTTGGCCAGGCTGTGTAACCTCCTTACCTCAAGTAGTCTGC 1215
 1045 eHsHisValGlyGlnAlaCysLeuGlnLeuLeuThrSerGlyGlyLeuL 1062
 1214 CTGCTCC. GCTTCCCAAGTGTGGGATTACAGGATGAGCCACCTGCC 1166
 1062 euAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHisAla 1078
 1165 CAG 1163
 1079 Arg 1079

seq_name: sp_human:Q99770

seq_documentation_block:
 ID Q99770 PRELIMINARY; PRT; 139 AA.
 AC Q99770;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE HYPOTHETICAL 15.4 KD PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]


```

RC      SEQUENCE FROM N.A.
RC      TISSUE-BRAIN:
RX      MEDLINE; 96207227.
RA      ANDERSON B., WENTLAND M.A., RICARRENTE J.Y., LIU W., GIBBS R.A.;
RT      "A "double adaptor" method for improved shotgun library
RT      construction.";
RL      Anal. Biochem. 236:107-113.(1996).
RP      [2]
RP      SEQUENCE FROM N.A.
RC      TISSUE-BRAIN:
RA      YU W., ANDERSSON B., WORLEY K.C., MUZNY D.M., DING Y., LIU W.,
RA      CAFFEYENTE J.Y., WENTLAND M.A., LENNON G., GIBBS R.A.;
RL      Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
RW      EMBL: U79260; AAB50206.1;
RW      Hypothetical protein.
SO      SEQUENCE 139 AA; 15443 MW; 47FA63F2 CRC32;

alignment_scores:
    Quality: 270.50      Length: 94
    Ratio: 3.559      Gaps: 2
    Percent Similarity: 80.851      Percent Identity: 62.766

alignment_block:
US-09-030-606-174/rev x Q99770  ..

Align seg 1/1 to: Q99770 from: 1 to: 139

1459 TTTTNTTTTTTTTTTTTGTAGACAGAGCTTACTGTGTGCCCCAGCTGG 1410
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
48 PheLeuPhePhePhePheGluThrGlnSerHisSerValThr.ArgLeuG 64
1409 AGTATAGGTGTGATGTCTCAACTTCACTTCAACCTGTGCTCCCATTTGA 1360
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
64 LucSerGlyThrIleSerIleHisCysAsnLeuGlySerProGlySer 80
1359 AGCATTTTCTCTGCTGAGCTGAGCTCCCAAGTAGTGGATTCAGAGGCGCTG 1310
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
81 SerAsnSerProAlaSerAlaSerAlaSerArgValAlaGlyThrAlaGlyThy 97
1309 CCACCATATCCAGCTAACTTTGTATTATTATTTAGTACAGACAGATTTCACC 1260
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
97 SarGArgArgGlnGlnLeuIlePheValPheLeuAlaGlnMetGlyPheHis 114
1259 ATTGTGGCCAGCGCTGCTGTGAATCTCTTACCTTCAAGTATGCTGCTGCC 1210
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||
114 IValaGlyArgAspGlyLeuAsnSprLeuAsnLeuValIleHisProProArg 130
1209 TCGCCTCCCAAGTGTGGGATTTTACAGCA 1180
||| ||||:||||:||||:||||:||||:||||:||||:||||:||||
131 Ser...ProlYsAlaLeuGlnGlyLeuGlnAla 139

seq_name: sp_human:O60448

seq_documentation_block:
ID O60448      PRELIMINARY;      PRT: 375 AA.
AC O60448:
DT 01-AUG-1998 (TEMBLrel. 07, Created)
DT 01-AUG-1998 (TEMBLrel. 07, last sequence update)
DT 01-AUG-1998 (TEMBLrel. 07, last annotation update)
DE NEURONAL THREAD PROTEIN ADTC-NTP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-NEURONAL;
RA DE LA MONTE S.M., GHANBARI K., FREY W., BEHESHTI I., HAUSER S.A.;
RA GHANBARI H.A., WANDS J.R.;
RL Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.
RW EMBL: AF010144; AAC06737.1;
SO SEQUENCE 375 AA; 41720 MW; 144FC2A2 CRC32;

```

```

Alignment_scores:
    Quality: 268.00      Length: 79
    Ratio: 4.254        Gaps: 0
    Percent Similarity: 79.747      Percent Identity: 64.557

alignment_block:
US-09-030-606-174/rev x 060448 ..

Align seg 1/1 to: 060448 from: 1 to: 375

1459 TTTTGTGAGACAGACTCTTGTCGGCCAGCTCG 1410
||||| ..... |||||..... |||||
297 PheanPheCysLeuPheGluMetIuSerHisSerValTrpGlnAlaGl 313
||||| ..... |||||..... |||||

1409 AGTAATAGAGCTGCATCTCAACTCAACCTCGCTGCCCATATTCA 1360
||||| ..... |||||..... |||||
313 yalGIInTProAsnLeuGlySerLeuGlnProLeuProIleuL 330
||||| ..... |||||..... |||||

1359 ACACAATCTCTGCTCCAGCTCCCAAGTAGCTGGATTACAGCGCTG 1310
||||| ..... |||||..... |||||
330 yaGrPheSerCysLeuSerLeuProSerSerTrpAspTyrGlyHisLeu 346
||||| ..... |||||..... |||||

1309 CCACATATCCAGCTAACTTTGTATTTTGATACAGACAGATTACCC 1260
||||| ..... |||||..... |||||
347 ProPheHisProIleAsnPhenCysLeuPheIleArgGlyGlyValSerPr 363
||||| ..... |||||..... |||||

1259 ATTTGGCCAGGCTGCTGTCAACCTCTTACCCTCAAG 1223
||||| ..... |||||..... |||||
363 oTyrlLeuSerGlyTrpSerGlnThrProAspLeuArg 375

seq_name: sp_human:Q15662

seq_documentation_block:
ID Q15662 PRELIMINARY; PRT; 368 AA.
AC Q15662:
DT 01-NOV-1996 (TEMBLrel_01, Created)
DT 01-NOV-1996 (TEMBLrel_01, Last sequence update)
DE 01-NOV-1996 (TEMBLrel_01, Last annotation update)
DS TRANSFORMATION-RELATED PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EPITHELIUM;
RA SHEN H., STEINBERG M.L.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DDJB databases.
DR EMBL: L24521; AAA36776.1; -.
FT NOY_TER 1 1
SQ SEQUENCE 368 AA; 42029 MW; E4772637 CRC32;

alignment_scores:
    Quality: 257.00      Length: 434
    Ratio: 1.332        Gaps: 18
    Percent Similarity: 44.470      Percent Identity: 26.728

Percent Similarity: 44.470      Percent Identity: 26.728

alignment_block:
US-09-030-606-174/rev x Q15662 ..

Align seg 1/1 to: Q15662 from: 1 to: 368

1440 AGACAGAGCTTACTCTGTGGCCCCAGCTGGAGTATAGTGTGATCTC 1391
||||| ..... |||||..... |||||
2 ArgLeuSerLeu...LeuSerProArgLeuGlnCysAnGlyMetIleLe 17

1390 AACTGACTGAACGCTCTGCTCCCATTA_TTGAAGAATCTTCGCGCCA 1342
||||| ..... |||||..... |||||
17 uAlaHisCysIlysLeuAlaGlnProGlyPheLysArgPheSerCysLeuS 34

1341 GCCTCCCAAGTAGCTGGGATTACAGAGCGCCCTGCCAACCATATCCAGTTAAC 1292

```

```

34  eileuProSeSerTrPaRyUgHnHsValProVoLg. glnValnI 50
1291  TTTTGATTTTAAAGTCAACAAGAGATTTACATTTTGGCAGGCTGCT 1242
      |||||
50  sPrEaUlpHeSeValGltThcGlyPHeHsArgLaGlaGlnAlaGyl 67
1241  TTAACATCCCTAACCAAGTGAATGCCGCGCTGCACCTC...CCAAATG 1195
      |||||
67  euGleuLeuHnHrSeSerValProProthnSer. AlaPheProLysCy 83
1194  CTGGGATTCACAGGCATAGACCCACCCCTGCCAACCCCTTTGTAGACTCTG 1145
      |||||
83  sTrPaRyUgHnHrSeSerValProProthnSer. AlaPheProLysCy 96
1144  ATTTTCCCTGTTCTCTGCTGGTTTCACCTCTATGAACnGCTGTCACGT 1095
      |||||
97  ....SerSerPheArgGlyLeuAlaHnLysPheAlaPhe 107
1094  TTCCCTCTGGGTACACAGTTCAGCCCTGAACAAC... 1061
      :|||
108  LeuProLys. PhePheAlaHnSProlleSeGlnPheGlnArgValGluC 124
1061  .... 1061
124  yasnValGlyCySProlleuLeuAlaMetLysUgLeuAlaUgSer 140
1060  ....CTGGTTTGAATGCACACAGTCCACCTTAATCTGGATTTTTCGA 1017
      |||||
141  SerLeuProGlyAlaSPheHnHrLeuUgYrPheUgYr... 153
1016  TAAACACATCAGAAAATTTTGAAGAGCTGCACACATTTAAAAAATTC 967
153  .... 153
966  ACAGACGAACCTGTAGCTTAGAAATATCAAAAAATTCACAAAGATAT 917
      :|||
154  ....TyrGlnGlnAla 158
916  ATCATGAATGCATAAACATATGCATAAATGCATATTAATGTTATGCT 867
      :|||
159  SerLeuAlaValLysasnLeuCyArgGlnPheHnHrSProlleUgYr 175
866  AAGGCTCCCCGTCACAAAGATATTTCTATGCACAGTTTGGGGAGCT 817
      :|||
175  rGlnLleSerHnLsLeuUgArgGlyValLleVal... 186
816  AAAGTATTAAGACGATTTATTTATGGGGTGGAGTGCACCGCCCTAGGCG 767
      :|||
187  ....AspAsnPheLeuLeu...HnSProlsPoly 195
766  CCAATGCTGGGAAGGTCAACATGTCATATGATGCTGTCTGCTGCT 717
      |||
196  ArgPheThTrpHrHle... 201
716  CCCCAGTCTGTCTGTTCTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 679
202  ....PhePheLeuSerTrValLysGlnAsn 211
678  CTATATTCGATGTTATTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCT 629
      |||||
211  eTrLeuValSPheHnHrPheHnGlyThGlnSerArgSerValAlaLeuLeu 227
628  GTACATGCTCTCCCTGTGTCTCTGCTCATCT...CTGCATCTGCCATGCG 582
      :|||
228  ProArgLeuGlnLysSerGlnAlaMetSerThrLleHnHnHnHnValLeuArg 244
581  GCCC... 578
244  gProAlaUgYrSerHnLsLleUgHnHrSProlsArgValLysGlnLysThnHnSP 261
578  .... 578
261  heLeuGlnValPheAlaHnLysArgGlnLysLeuGlnHnLysLleLeuLys 277

```

```

577 ..... TGAGTGTCTGCTGTCCTCCGCTTCCACCTCC 563
278 ThrProAsnProLeucylsAlaLeuHisSerAlaProSerProSerLeuPr 294
542 TGAG.....TCGCTGCTCCCTCCCTTCCGCTCT..... 512
| ..... ||||| ||||| |||||
294 oProPheLeuArgCysThrGlyArgLeuProPheTyLeuGlyLeuAspA 311
511 .....CAGCTGCAGAGACTCTTGTGCTTGTGCTCCCTCCGCGG.... 474
| ||||| ||||| ||||| |||||
311 sPheLeuPheValAlaGlyAlaLeuMetPheLeuProValSerPheLeu 327
473 .....CGGAGACATGNNGGGGGTGTACGCGGCAATAGACCTTANN 430
| ..... ||||| ||||| |||||
328 AsnProHisThrLeuThrTrpProProGlnCysCysThrArgSerAsp 344
429 CAGACCTCMTACAGACACCGACGCTTCACGACTGACAGCAC 386
| ..... ||||| ||||| ||||| |||||
344 sAsnProLeuArgGlyGlnArgGluIleSerAlaLeuSerHis 358

seq_name: sp_invertebrate:Q26195

seq_documentation_block:
ID Q26195 PRELIMINARY; PRT; 200 AA.
AC Q26195;
DT 01-NOV-1996 (TREMblrel. 01. Created)
DT 01-NOV-1996 (TREMblrel. 01. Last sequence update)
DI 01-NOV-1998 (TREMblrel. 08. Last annotation update)
DE PVAL GENE.
GN PVAL.
OS Plasmodium vivax.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-INDIAN;
RX MEDLINE; 98171310.
RA DHAR A., GUPTA S., SHARMA Y.D.;
RT "Alu elements in a Plasmodium vivax antigen gene.";
RL FEBS Lett. 423:193-197(1998).
DR EMBL; X92485; CAA633219.1; -.
SQ SEQUENCE 200 AA: 22833 MW: 8233BC4F CRC32;

alignment_scores:
Quality: 247.00 Length: 94
Ratio: 3.293 Gaps: 1
Percent Similarity: 79.787 Percent Identity: 63.830

alignment_block:
US-09-0330-606-174/rev x Q26195 ..

Align seg 1/1 to: Q26195 from: 1 to: 200

1457 TTTTGTGTTTGTGTTGAGACAGAGTCTTACTGTGGCCGACGTGAG 1408
| ||||| ||||| ||||| ||||| ||||| ||||| |||||
108 PheIlePhePhePhePheArgPheValLeuLeuCysHisProGlyTrpSe 124
1407 TAAAGGCGTGTATGTCACACTGCTCAACGCTGGCCGCTCCCAATTCAG 1358
| ||||| ||||| ||||| ||||| ||||| |||||
124 TAlaValAlaGlnSerLeuPheThrValAlaSerThrPheLeuValLys 140
1357 CAATTTCCTGCTCAGCCTCCCAAGTAGCTGGATTACAGCGGCTGCC 1308
| ||||| ||||| ||||| ||||| ||||| ||||| |||||
141 GlnSerCysLeuGlyLeuProSerSerThrPAspTyrAlaArgTyleTr 157
1307 ACCAATTCAGCTAACTTTTGATTTTGAATACAGACAGAGAT TTACACA 1259
| ||||| ||||| ||||| ||||| ||||| ||||| |||||
157 oProHisLeuAlaAsnPheSerPhePheCysArgAsnLysSerLeuThrM 174
1258 TTTTGGCAGAGCGTGTGGAATCTCTTACCTTAACATAATGATCTCGCT 1209
| ..... ||||| ||||| ||||| ||||| ||||| |||||
174 etLeuProAlaGlyLeuIleLeuAsnSerTrpProIleValIleLeuPro 190

```

DR EMBL; X82223; CAA57701.1; -

KW Signal: Hydrolase.
 FT SIGNAL 1 13 POTENTIAL.
 DT CHAIN 14 242 PROTRYSIN.
 SO SEQUENCE 242 AA: 26201 MW: DDD27DCA CRC32:

alignment_scores:
 Quality: 221.00 Length: 216
 Ratio: 1.826 Gaps: 10
 Percent Similarity: 56.019 Percent Identity: 30.093

alignment_block:

US-09-030-606-174 x 092099 ..

Align seg 1/1 to: 092099 from: 1 to: 242

```

3 TCAGCCGACACTGTTCCAGAGAGTGTGAGAGCTCTACACCATCGG 52
|||||
57 SerAlaAlaHisCysTyrLysSerArgValu.....ValAr 69
GCTGGCCCTGACAGTCTTGAGCCGACCAAGAGGAGGAGGAGCATGG 102
|||||
69 GMeGlyGlnHisHisIleArgVal...ThGlnGlyLysGlnInphei 85
103 TGAAGCCAGCCTCTCCGTACGACCCAGAGTACACAGACCTTGCTC 152
|||||
85 leSerSerArgValIleArgHisProAsnTyrSerTyrAsnIle 101
153 GCTACAGCCTCATGCTCATCAAGTGTGACCAATCCGTCGAGTCTGA 202
|||||
102 AspaSnpSPrlMetLeuIleLysLeuSerLysProAlaTrLeuAsnGI 118
203 CACATCCGAGAGCATGATGCTTGTGAGTGCCTTACCGCGGGAAGT 252
|||||
118 nTyrValGlnAlaValAlaLeuProSerSerCysAlaProAlaGlyTm 135
253 CTTCGCTGTTCTGCTGGGCTGTGCTGCGAAAGGTGAGCTCAGGGT 302
|||||
135 etCysThrValSerGlyTyrGly..... 142
303 GTGTGTGCTGCTTCAAGAGAGTCTCTGACAGTGCAGGCGGAGTAC 352
|||||
143 .....SerThrInSer..... 146
353 CCAGAGCTGCTGCTCCAGAGAGATGCTTACCGTGTGAGTGTGA 402
|||||
147 ....SerSerAlaasp..GlyAsnLys.....LeuGlnCysLeuAs 158
403 CGTGTGCTGCTGTCTGANGAGGTCTGCANTAACTCTATGACCGCTGT 452
|||||
158 nIleProIleLeuSerAspArgAspCysAspAsnSerTyrProGlyMetI 175
453 ACCACCCANCATGTTCTGCGCCGCGGAGGAGCAAGACCAAGAGACTCC 502
|||||
175 leThrAspAlaMetPheCysAlaGlyTyrLeuGlnGlyLysAspSer 191
503 TGCAACGTGAGAGAGGAGGAGGAGGAGGAGGAGCTCAGAGGAGG.. 550
|||||
192 Cys.....GlnGlyAspSerGlyGlyProValValCysAsnGlyGI 205
551 .....TGAGAGAGGGG.....AGACAGAGACACA 575
|||||
205 uLeuGlnGlyValValSerTrpGlyTyrGlyCysAlaGlnArgSpHSP 222
576 CAGGCGCCGATGGCGAGATG.....CAGAGTGTGAGAGACACA 613
|||||
222 roGlyValValTyrAlaLysValCysLeuPheAsnAspTrpLeuGlnThr 237

```

seq_name: sp-vertebrate:093265

seq_documentation_block:
 ID 093265 PRELIMINARY; PRT; 250 AA.
 AC 093265;

DT 01-NOV-1998 (TREMblrel. 08, Created)
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
 DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)
 DE TRYPsinogen 1 PRECURSOR (EC 3.4.21.4).
 GN TRP1.
 OS Pleuronectes americanus.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
 OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
 OC Pleuronectiformes; Pleuronectoidae; Pleuronectidae; Pleuronectes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PANCREAS;
 RA DOUGLAS S.E., GALLANT J.W.;
 RT Isolation of cDNAs for Trypsinogen from the Winter Flounder,
 RT Pleuronectes americanus.";
 RL J. Mar. Biotechnol. 0:0-0(1998).
 DR EMBL; AF012462; AAC32751.1; -.
 DR PFM; PF00089; trypsin; 1.
 KW Signal; Hydrolase.
 FT SIGNAL 1 15 POTENTIAL.
 FT CHAIN 23 250 TRYPsinogen 1.
 SO SEQUENCE 250 AA: 27466 MW: D387394D CRC32:

alignment_scores:
 Quality: 220.00 Length: 192
 Ratio: 2.018 Gaps: 7
 Percent Similarity: 56.771 Percent Identity: 31.771

alignment_block:

US-09-030-606-174 x 093265 ..

Align seg 1/1 to: 093265 from: 1 to: 250

```

3 TCAGCCGACACTGTTCCAGAGAGTGTGAGAGCTCTACACCATCGG 52
|||||
59 SerValAlaHisCysTyrTrpAsn.....ProTyrAlaMetGlnValme 73
53 GCTGGCCCTGACAGTCTTGAGCCGACCAAGAGGAGGAGGAGCATGG 102
|||||
73 tLeuGlyGlnHisAsnLeuArgVal...PheGlnGlyThrGlnGlnLeu 89
103 TGAAGCCAGCCTCTCCGTACGACCCAGAGTACACAGACCTTGCTC 152
|||||
89 etLysThrAspTrpIleIleTrpHisProAsnTyrAspTrpGlnThrLeu 105
153 GCTACAGCCTCATGCTCATCAAGTGTGACGAATCCGTCGAGTCTGA 202
|||||
106 AspPheAspTrpIleMetLeuIleLysLeuTyrHisProValGlnValThrGI 122
203 CACATCCGAGAGCATGACATGCTTGTGAGTGCCTTACCGCGGGAAGT 252
|||||
122 uAlaValAlaProIleSerLeuProSerSerCysProValGlyGlyThrP 139
253 CTTCGCTGTTCTGCTGGGCTGTGCTGCGAAAGGTGAGTCAAGGGT 302
|||||
139 roCysSerValSerGlyTyrGly..... 146
303 GTGTGTGCTGCTTCAAGAGAGTCTCTGACAGTGCAGGCGGAGTAC 352
|||||
147 .....AsnThrAlaArgasp..GlyAspAsp 154
353 CCAGAGCTGCTGCTCCAGAGAGATGCTTACCGTGTGAGTGTGA 402
|||||
155 Val.....TyrMetProThrLeuGlnInpHSP 165
403 CGTGTGCTGCTGTCTGANGAGGTCTGCANTAACTCTATGACCGCTGT 452
|||||
165 pValProIleIleAspGlnGlnGlnCysMetLysSerTyrProGlyMetI 182
453 ACCACCCANCATGTTCTGCGCCGCGGAGGAGGAGCAAGACCAAGAGACTCC 502
|||||
182 leSerProArgMetValCysAlaGlyPheMetAspGlySerArgAspAla 198

```


FT NON_TER 1 1
FT NON_TER 178 178
SQ SEQUENCE 178 AA: 19352 MW: 506224A9 CRC32:

alignment_scores:
Quality: 212.50 Length: 202
Ratio: 1.986 Gaps: 6
Percent Similarity: 52.970 Percent Identity: 27.723

alignment_block:

US-09-030-606-174 x 093594 ..

Align seg 1/1 to: 093594 from: 1 to: 178

```
3 TCAGCCGACACGTTCAGAACTGAGTGCAGAGCTCTACACATCGG 52
|||||
23 SerAlaIaHisCysTyrLysSerArgValGlu.....ValAr 35
53 GCTGGGCTCCACAGTCTTGAGCCGACCAAGAGCCAGAGCCAGATGG 102
|||||
35 GluGlyGlnHisAsnIleArgVal...ThrGluAsnThrGluInPheI 51
103 TGGAGCCACCTCTCCGACGCGACCCAGAGTACAAACAGACCTTGCTC 152
|||||
51 LeuSerSerArgValIleArgHisProArgTyrSerSerTyrAsnIle 67
153 GCTAACGACCTCATGCTCATCACTTGAGCAATCCGTGCCAGTCTGA 202
|||||
68 AspAsnAspIleMetIleLeuLeuLeuSerSerProAlaThrLeuAsnG 84
203 CACCATCCGAGCATGACATTCCTGCAAGTCCCTACCGCGGGAAC 252
|||||
84 nIyRValGlnProValAlaLeuProThrSerCysAlaProIaGlyThrM 101
253 CTGCTCTGTTCTGCTGGGGGCTGCTGGCGAAGGTGAGCTCAGCGGT 302
|||||
101 eTcysThrValSerGlyTyrpGly..... 108
303 GTGTGTCTGCCCTTTCAGAGAGGTCTCTGCCAGTCGCGGGGGCTGAC 352
108 ..... 108
353 CCAAGACCTTGCGTCCAGGCGAGATGCTTACGCTGCGAGTGTGA 402
|||||
109 .AsnThrMetSerSerThrAlaAsp..ArgAsnLysLeuGlnCysLeuAs 124
403 CGTGTGCGGTGTCTGANGAGGTCTGCANTATGCTCTATGACCGCTGT 452
|||||
124 nIleProIleLeuSerPheLysAspCysAspAsnSerTyrProGlyMetI 141
453 ACCAACCACATGTCTGCGCCGCGGAGGCGAAGACCAAGAGACTCC 502
|||||
141 leThrAspAlaMetPheCysIaGlyTyrLeuGlnGlyGlyAspSer 157
503 TGCAACTGAGAGAGGGAAGAGGAGGCGAGCTCAGGGAAGGTTG 552
|||||
158 CysGlnInLysPserGly..... 163
553 GAGAAGGGGAGAGACAGACACAGGCGCG.....CATGCGAGA 593
|||||
164 .....GlyProValValCysAsnGlyGlnL 172
594 TGCAG 598
172 eugIn 173
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:29:36 ; Search time 1809.22 seconds
(without alignments)
1590.700 Million cell updates/sec

Title: US-09-030-606-174
Perfect score: 1459
Sequence: 1 GGTCCAGCCGACACACTGTTC.....TCAAAAAAAAAAAAAAAAAAAAAA 1459

Scoring table: IDENTITY_NNC

Searched: 2546578 seqs, 98626752 residues

Database :

EST:
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: em_est20:*
21: em_est21:*
22: em_est22:*
23: em_est23:*
24: em_est24:*
25: em_est25:*
26: em_est26:*
27: em_est27:*
28: em_est28:*
29: em_est29:*
30: em_est30:*
31: em_est31:*
32: em_est32:*
33: em_est33:*
34: em_est34:*
35: em_est35:*
36: em_est36:*
37: em_est37:*
38: em_est38:*
39: em_est39:*
40: em_est40:*
41: em_est41:*
42: em_est42:*
43: em_est43:*
44: em_est44:*
45: em_est45:*
46: em_est46:*
47: em_est47:*
48: em_est48:*
49: em_est49:*
50: em_est50:*
51: em_est51:*
52: em_est52:*
53: em_est53:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	358.4	24.6	360	50	A1675523	A1675523 wc01f01.x
2	286.4	19.6	777	48	A1557281	A1557281 PT2.1.15
3	245.4	16.8	569	50	A1686689	A1686689 t035g11.x
4	244.6	16.8	415	35	AA551449	AA551449 nj55e05.s
5	242.8	16.6	722	48	A1557025	A1557025 PT2.1.10
6	226.2	15.5	481	50	A1669421	A1669421 ty32d03.x
7	225.4	15.4	415	48	A1609972	A1609972 t078e10.x
8	223	15.3	407	35	AA535216	AA535216 nj75f05.s
9	222.4	15.2	479	26	W96522	W96522 ze43f08.r1
10	222.2	15.2	437	36	AA644090	AA644090 ab52b04.s
11	220.2	15.1	500	36	AA626040	AA626040 af94a08.s
12	220.2	15.1	454	49	A1634187	A1634187 ts55a08.x
13	220	15.1	374	38	AA746911	AA746911 nx69c10.s
14	219.8	15.1	434	36	AA630854	AA630854 nt57f09.s
15	219.6	15.1	388	39	AA837686	AA837686 oe05h09.s
16	219.4	15.0	323	35	AA536040	AA536040 nj81c09.s
17	219.2	15.0	484	33	AA410788	AA410788 zt35b11.r
18	219.2	15.0	380	36	AA613624	AA613624 nol8d03.s
19	218.4	15.0	458	51	A1733856	A1733856 z019c03.y
20	218.2	15.0	408	34	AA486877	AA486877 ab16h04.s
21	218	14.9	533	39	AA833875	AA833875 od64e08.s
22	217.2	14.9	470	34	AA456924	AA456924 aa90b09.s
23	217.2	14.9	329	36	AA643770	AA643770 np06e10.s
24	217.2	14.9	466	42	A1087040	A1087040 oy70a10.x
25	217.2	14.9	365	46	A1419337	A1419337 tf27h01.x
26	217.2	14.9	397	46	A1421950	A1421950 lf45d05.x
27	217	14.9	415	34	AA515048	AA515048 ng67h10.s
28	216.6	14.8	461	30	AA225406	AA225406 nc24d02.r
29	216.6	14.8	301	39	AA828592	AA828592 od74e10.s
30	216.4	14.8	528	39	AA833896	AA833896 od64g08.s
31	216.4	14.8	344	41	A1054030	A1054030 q166a12.x
32	216.2	14.8	428	41	A1066646	A1066646 oz82d10.x
33	216	14.8	569	53	HSW007223	A1042373 Homo sapi
34	215.6	14.8	367	45	A1361090	A1361090 gy04d06.x
35	215.6	14.8	524	47	A1524240	A1524240 th11d12.x
36	215.4	14.8	457	31	AA284247	AA284247 zc65d01.T
37	215.4	14.8	480	49	A1635028	A1635028 tz03d06.x
38	215.2	14.7	376	29	AA176978	AA176978 nc01c01.s
39	215.2	14.7	474	46	A1457313	A1457313 tl73a05.x
40	215	14.7	440	36	AA601278	AA601278 nol5b09.s
41	214.8	14.7	624	36	AA601356	AA601356 nol6b05.s
42	214.8	14.7	623	33	HSW003412	A1038936 Homo sapi
43	214.6	14.7	441	29	AA169245	AA169245 zp19d03.s
44	214	14.7	357	45	A1340832	A1340832 lb55a10.x
45	214	14.7	413	46	A1433104	A1433104 th41d12.x

ALIGNMENTS

RESULT 1
A1675523 360 bp mRNA EST 19-MAY-1999
wc01f01.x1 NCI_CGAP_PT28 Homo sapiens cDNA clone IMAGE:2313913 3',
ACCESSION A1675523
NID 94876003
VERSION A1675523.1 GI:4876003

QY 385 CGCTGCTGACGCTGAGACGTGCTGCTGTCTGANGAGCTTCANTAGCTCTATGA 444
 |||||||
 Db 387 CGTCTGTAGAGTGGTGAACGTGTGCTGTCTGAGAGAGTCTGACACTAAGCTCTATGA 446
 |||||||
 QY 445 CCCGCTGTACCAACCCCAACATGTTCTGCGCGCGGAGGAGGAGCAAGACAGAGACTCTG 504
 |||||||
 Db 447 CCCCTGTACCAACCCCAACATGTTCTGCGCGCGGAGGAGGAGCAAGACAGAGACTCTG 506
 |||||||
 QY 505 CAACG 509
 |||||||
 Db 507 CAACG 511

RESULT 3
 A1686689/c
 LOCUS A1686689 569 bp mRNA EST 27-MAY-1999
 DEFINITION lu35911.x1 NCI_CGAP_Pr28 Homo sapiens CDNA clone IMAGE:2253092 3' similar to TR:092046 Q92046 PREPROTRYPsin PRECURSOR ;, mRNA sequence.
 ACCESSION A1686689
 NID 94897983
 VERSION A1686689.1 GI:4897983
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 569)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3189584.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bdrip/image/image.html

Seq primer: -40UP from Gibco
 High quality sequence stop: 444.
 Location/Qualifiers

FEATURES
 source

1..569
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2253092"
 /clone_id="NCI_CGAP_Pr28"
 /sex="male"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Pr28 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 985608-986759, 1101192-1101959, and 1217928-1220615)."
 Subtraction by Bento Soares and M. Fatima Bonaldo.
 BASE COUNT 125 a 150 c 170 g 124 t

Query Match 16.8%; Score 245.4; DB 50; Length 569;
 Best Local Similarity 79.6%; Pred. No. 2.3e-36;

Matches 339; Conservative 0; Mismatches 4; Indels 83; Gaps 1;
 QY 84 GAGCAGGAGGAGCCAGATGGTGGAGGCCAGGCTCTCCCTAGGACCCAGAGTCAACAGA 143
 |||||||
 Db 569 GAGCAGGAGGAGCCAGATGGTGGAGGCCAGGCTCTCCCTAGGACCCAGAGTCAACAGA 510
 |||||||
 QY 144 CCGTGTGCTGCTAAAGCACTCATGCTCATCAAGTTGGAGCAATCCGTGTCGAGTCTGAC 203
 |||||||
 Db 509 CCGTGTGCTGCTAAAGCACTCATGCTCATCAAGTTGGAGCAATCCGTGTCGAGTCTGAC 480
 |||||||
 QY 204 ACCATCCGAGACATCAAGTCTGCTGCAAGTCCCTAACCGGGGAACTTGTCCCTGTT 263
 |||||||
 Db 449 ACCATCCGAGACATCAAGTCTGCTGCAAGTCCCTAACCGGGGAACTTGTCCCTGTT 390
 |||||||
 QY 264 TCTGGCTGGGGTCTGCTGGGGAACGGTGAAGTCAAGGGTGTGTCTGCTTCAAGG 323
 |||||||
 Db 389 TCTGGCTGGGGTCTGCTGGGGAACGGTGAAGTCAAGGGTGTGTCTGCTTCAAGG 365
 |||||||
 QY 324 AGGTCTCTGCCAGTGGCGGGGCTGACCAAGAGCTCTGCTCCAGAGAGTGCCTA 383
 |||||||
 Db 365 -----GGCAGATGCTTA 353
 |||||||
 QY 384 CCGTGTGACAGTGGTGAACGTGCTGCTGTCTGANGAGTCTGCANTAGCTCTATG 443
 |||||||
 Db 352 CCGTGTGACAGTGGTGAACGTGCTGCTGTCTGANGAGTCTGCANTAGCTCTATG 293
 |||||||
 QY 444 ACCGCTGTACCAACCCCAACATGTTCTGCGCGCGGAGGAGCAAGACAGAGACTCT 503
 |||||||
 Db 292 ACCGCTGTACCAACCCCAACATGTTTTCGCGCGGAGGAGCAAGACAGAGACTCT 233
 |||||||
 QY 504 GCAACG 509
 |||||||
 Db 232 GCAACG 227

RESULT 4
 AA551449
 LOCUS AA551449 415 bp mRNA EST 05-SEP-1997
 DEFINITION nj55605.s1 NCI_CGAP_Pr9 Homo sapiens CDNA clone IMAGE:996416 similar to SW:RLX_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR ;, mRNA sequence.

ACCESSION AA551449
 NID 92321701
 VERSION AA551449.1 GI:2321701
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 415)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 9, 1995 this sequence version replaced gi:802243.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuquib, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: David B. Kitzman, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bdrip/image/image.html

Insert Length: 640 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amerham
 High quality sequence stop: 412.
 Location/Qualifiers
 1..415

FEATURES
 source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1ib="IMAGE:996416"
/clone_1ib="NCI_CGAP_Prg"
/sex="male"
/tissue_type="normal prostatic epithelial cells"
/lab_host="DH10B"
/note="Organ: prostate; Vector: PAMP10; mRNA made from
normal prostatic epithelial cells, cDNA made by oligo-dt
priming. Non-directionally cloned. Size selected on
agarose gel, average insert size 600 bp. Library made by
D. Krizman, NIH."
BASE COUNT      78 a 125 c 129 g      83 t
ORIGIN

```

```

Query Match      16.8%; Score 244.6; DB 35; Length 415;
Best Local Similarity 78.2%; Pred. No. 3.3e-36;
Matches 344; Conservative 0; Mismatches 12; Indels 84; Gaps 1;

```

```

QY 70 TTGAGCGCCGACAGAGCCAGGAGCCAGATGNGTGGAGGCGACGCTCTCGTACGGCACC 129
DB 2 TCGACGCGGACAGAGCCAGGAGCCAGAGCGTGGAGGCCAGCTCTCGTACGGCACC 61
QY 130 CAGAGTACAGAGAGCCCTTCTGCTGCTACGACCTCATGCTCATCAAGTTGGAGCAATCCG 189
DB 62 CAGAGTACAGAGAGCCCTTCTGCTGCTACGACCTCATGCTCATCAAGTTGGAGCAATCCG 121
QY 190 TGTCCGAGTGTACACCATCCGAGACATCAGCATTTGCTTGGAGTCCCTACCGGGGGA 249
DB 122 TGTCCGAGTGTACACCATCCGAGACATCAGCATTTGCTTGGAGTCCCTACCGGGGGA 181
QY 250 ACTCTGCTGCTTGTGGTGGGGTGTGCTGGGGAACGGTGAAGTCAACGGGTGTGTCTC 309
DB 182 ACTCTGCTGCTTGTGGTGGGGTGTGCTGGGGAACGGTGTGTGTGTGTGTGTGTGTGT 222
QY 310 TGCCCTCTTCAAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 369
DB 222 ----- 222
QY 370 AGGCAGATGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 429
DB 222 ----CAGATGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 277
QY 430 CANTAGCTGTATGAGCCGCTGTACACCCAGCCAGCATGTTGCGCGGGGAGGAGGAGAGA 489
DB 278 CAGTAACTGTATGAGCCGCTGTACACCCAGCCAGCATGTTGCGCGGGGAGGAGGAGAGA 337
QY 490 CCAGAGAGACTCTGCAACG 509
DB 338 CCAGAGAGACTCTGCAACG 357

```

```

RESULT 5
LOCUS AI557025 722 bp mRNA EST 23-MAR-1999
DEFINITION PT2.1.10.F05.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION AI557025
NID 94489388
VERSION AI557025.1 GI:4489388
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 722)
AUTHORS Huang,G.M., Ng,W., Farakas,J., Chen,L., Llang,H.A., Gordon,D., Jun
Yu,J. and Hood,L.
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138511.
Contact: Guyang Matthew Huang

```

```

Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of
Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.

```

```

FEATURES
SOURCE 1..722
Location/Qualifiers

```

```

BASE COUNT      146 a 156 c 195 g      176 t      49 others
ORIGIN

```

```

Query Match      16.6%; Score 242.8; DB 48; Length 722;
Best Local Similarity 90.0%; Pred. No. 6.7e-36;
Matches 298; Conservative 0; Mismatches 25; Indels 8; Gaps 4;

```

```

QY 36 AGCTCTACACATGGGCTGGGCTGCGACAGTCTTGAGGCGGACCAAGAGCCAGGAGGC 95
DB 115 AGCTTTACACATGGGCTGGGCTGCGACAGTCTTGAGGCGGACCAAGAGCCAGGAGGC 174
QY 96 CAGATGTTGAGAGCCAGCCCTTCCGTCACGCGCACCCAGATACAGAGACCTTGTGCT 155
DB 175 CAGATGTTGAGAGCCAGCCCTTCCGTCACGCGCACCCAGATACAGAGACCTTGTGCT 234
QY 156 AACGACCTATGCTCATCAAGTTGACGATCCGTGTCCGAGTGTGACCAATCCGAGAC 215
DB 235 AACGACCTATGCTCATCAAGTTGACGATCCGTGTCCGAGTGTGACCAATCCGAGAC 294
QY 216 ATGAGCATGCTTGCAGAGGCGCTACCGGGGGAACCTGTGCC-TGCTTCTGGGCTGGGG 274
DB 295 ATGAGCATGCTTGCAGAGGCGCTACCGGGGGAACCTGTGCC-TGCTTCTGGGCTGGGG 354
QY 275 TCTGC-TGGCGAAGGCTGAGTCTACAGGAGTGTGCTGCTGCTCTTC---AAGGAGGTCC 329
DB 355 TCTGTTGGGGAAGGCTGAGTCTACAGGAGTGTGCTGCTGCTCTTC---AAGGAGGTCC 414
QY 330 TCTGCCCA--GTGCGGGGGCTGACCCAGAG 358
DB 415 TTTTNCMAAGTNTGCGGGGCTNACCAGAG 445

```

```

RESULT 6
LOCUS AI669421/c 481 bp mRNA EST 14-MAY-1999
DEFINITION ty32d03.x1 NCI_CGAP.Ut2 Homo sapiens cDNA clone IMAGE:2280773 3'
ACCESSION AI669421
NID 94834195
VERSION AI669421.1 GI:4834195
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 481)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1130773.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

```


RESULT	10
AA644090/c	
LOCUS	AA644090 437 bp mRNA
DEFINITION	ab022044.s1 StrataGene lung carcinoma 937218 Homo sapiens cDNA clone IMAGE:845559 3' similar to contains Alu repetitive element// mRNA sequence.
ACCESSION	AA644090
NID	92569308
VERSION	AA644090.1 GI:2569308
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estevenson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lln.gov) for further information.
Insert Length: 834 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 416.

FEATURES	SOURCE	LOCATION/Qualifiers
		1..437
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone_image:843539"
		/clone_lib="Stratagene lung carcinoma 937218"
		/tissue_type="lung carcinoma"
		/cell_line="NCI-H69"
		/dev_stage="cell line NCI-H69"
		/lab_host="SOLR (kanamycin resistant)"
		/note="Organ: lung; Vector: pBluescript SK-; Site: 1; EcoRI; Site: 2; XhoI; Cloned unidirectionally. Primer: Oligo dT. Small cell carcinoma cell line NCI-H69. Average insert size: 1.0 kb, Uni-ZAP-XR Vector; -5' adaptor sequence: 5' CTCAGCTTTTTTTTTTTTGGGAC 3' -3' adaptor sequence: 5' CTCAGCTTTTTTTTTTTTGGGAC 3'."
BASE COUNT		98 a 112 c 78 g 149 t
ORIGIN		

RESULT	11			
LOCUS	AA626040/c			
DEFINITION	AA626040	500 bp	mRNA	EST 15-OCT-1997
ACCSSION	A194a08:1	Soares_Totai	tetus.Nb2HF8_9w	Homo sapiens CDNA clone
NID	IMAGE:1055414	3'	similar to	contains Alu repetitive
VERSION	AA626040	element:contains	element MERR27	repetitive element ;, mRNA sequence.
KEYWORDS	G2538427			
SOURCE	AA626040.1	GI:2538427		
ORGANISM	human.			
	Homo sapiens			

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 500)
Hallier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Kizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
Mattlin,J., Moore,B., Schellenberg,K., Stepoe,M., Tan,F.,
Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
On Apr 14, 1993 this sequence version replaced g1:503350.

CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
 This clone is available royalty-free through LINT ; contact the
 IMAGE Consortium (infoimage.lln.gov) for further information.
Seq Primer: -40ml3 fwd, ET from Amersham.
Location/Qualifiers
 1..500

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1055414"
/clone_id="Soares_total_fetus_Nb2Hf8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/notes="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'

```

TGTACCAATCTGAAGTGGAGCGGCCCTTAATTTTTTTTTTTT 3'}.
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pTZ19 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo. "
 BASE COUNT 115 a 123 c 107 g 154 t 1 others
 ORIGIN

Query Match 15.1%; Score 220.2; DB 36; Length 500;
 Best Local Similarity 81.2%; Pred. No. 1e-31; Indels 0; Gaps 0;
 Matches 253; Conservative 0; Mismatches 59;

QY 1140 AATATCAAGACTCTTCAAGAGAGCTGGGAGGCTGCTATCTTAATCCAGACATT 1199
 DB 483 AAGATTAAAGAACTTCCCTCCGNCAGGGGTGTGTTCACTCTTAATCCAGACATT 424
 QY 1200 TGGGAGCGCAGGCGAGCATCTTGAAGTAAGAGTTCAAGACCGCTGGCCAAAT 1259
 DB 423 TCGGCGCGCAGGCGAGCATCTTGAAGTAAGAGTTCAAGACCGCTGGCCAAAT 364
 QY 1260 GGTGAATCTCTCTCTACTTAAATCAAAATCTGATGATGCTGGCAGCGCCCTG 1319
 DB 363 GGTGAATCTCTCTCTACTTAAATCAAAATCTGATGATGCTGGCAGCGCCCTG 304
 QY 1320 TAATCCAGACTCTTGAAGAGCTGGAGGAGTAATGCTTGAATGGAGCAGACAGT 1379
 DB 303 TAATCCAGACTCTTGAAGAGCTGGAGGAGTAATGCTTGAATGGAGCAGACAGT 244
 QY 1380 TGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1439
 DB 243 TGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 184
 QY 1440 TCAAAAAAATAA 1453
 DB 183 TCAAAAAAATAA 170

RESULT 12
 LOCUS A1634187/c 454 bp mRNA EST 26-APR-1999
 DEFINITION t535a08.x1 NCI-CGAP_K168 Homo sapiens cDNA clone IMAGE:2232470 3'
 similar to contains Alu repetitive element; contains element TAR1
 repetitive element; mRNA sequence.
 ACCESSION A1634187
 NID 94685517
 VERSION A1634187.1 GI:4685517
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 454)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 18, 1998 this sequence version replaced gi:3138239.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bhrp/image/image.html
 Seq primer: -40UP from Gibco

FEATURES
 source
 High quality sequence stop: 411.
 Location/Qualifiers
 1..454
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="5"
 /clone="IMAGE:2232470"
 /clone_id="NCI-CGAP_K168"
 /tissue_type="renal cell tumor"
 /lab_host="DH10B"
 /note="Organ: kidney; Vector: pCMV-SPORT6; Site-1: SalI;
 Site-2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.2 kb. Life Technologies catalog #: 11524-014"

BASE COUNT 118 a 103 c 93 g 140 t
 ORIGIN
 Query Match 15.1%; Score 220.2; DB 49; Length 454;
 Best Local Similarity 81.8%; Pred. No. 1e-31;
 Matches 266; Conservative 0; Mismatches 58; Indels 1; Gaps 1;

QY 1130 AATAACAGGAAAAATCAAGACTCTTCAAGAGCTGGCAGGCTGCTCATGCTGTAT 1189
 DB 325 AATAATTTATTTATTAATAAATACCAAAAGAGCTGGCCTGTGCTACAGCTGTAT 266
 QY 1190 CCGACACTTTGGAGG-CGAGGCGAGCAGATCACTTGAAGTAAGGATCAAGCCAGC 1248
 DB 265 CCGACACTTTGGAGGCGAGGCGAGTGTGATCACTTAAGTCAGGATTCAGACCCAGC 206
 QY 1249 CTGGCCAAATGCTGAATCTCTGTCTACTTAAATCAAAAGTTAGCTGATATGTG 1308
 DB 205 CTGGCCAAATGCTGAATCTCTGTCTACTTAAATCAAAAGTTAGCTGATATGTG 146
 QY 1309 CGAGGCGCTGTATCCAGCTACTTGGAGGCTGAGCAGAGCATTTGTTGAATATG 1368
 DB 145 GCGTGGGCTGTAGTCCAGCTACTCGGAGGCTGAAACAGAGATCGCTTGAACCCGG 86
 QY 1369 GAGGCGAGGTTGAATGAGTGTGATGATGATGATGATGATGATGATGATGATGAT 1428
 DB 85 GAGGCGAGGTTGAATGAGTGTGATGATGATGATGATGATGATGATGATGATGAT 26
 QY 1429 AAGACTGTGTCTCAAAAAAATAA 1453
 DB 25 GAGACTGTGTCTCAAAAAAATAA 1

RESULT 13
 LOCUS AA746911 374 bp mRNA EST 22-JAN-1998
 DEFINITION nx59c10.s1 NCI-CGAP_A1v1 Homo sapiens cDNA clone IMAGE:1267506
 similar to contains Alu repetitive element; mRNA sequence.
 ACCESSION AA746911
 NID 92786869
 VERSION AA746911.1 GI:2786869
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 374)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Dec 12, 1995 this sequence version replaced gi:1119179.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
 Ph.D.
 CDNA Library Preparation: David B. Kitzman, Ph.D.

AA837686

ACCESSION . AA837686
UTD -001000

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 23:25:55 ; Search time 2910.9 Seconds
(without alignments)
1594.033 Million cell updates/sec

Title: US-09-030-606-174
Perfect score: 1459
Sequence: 1 GGTACGCCGACACTGTTTC.....TCAAAAAAAAAAAAAAAAAAAA 1459

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenBank:*

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_om:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_v1:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_v1:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1411.4	96.7	4740	42	AF135023	AF135023 Homo sapi

2	1411.4	96.7	4385	42	AF148532	AF148532 Homo sapi
3	1405	96.3	5900	42	AF113141	AF113141 Homo sapi
4	296.4	20.3	1347	42	AF113140	AF113140 Homo sapi
5	248.2	17.0	166723	34	HS310013	AL031658 Homo sapi
6	240.6	16.5	151770	10	HSAC002066	AC002066 Homo sapi
7	240	16.4	166663	35	AC007728	AC007728 Homo sapi
8	238.8	16.4	1234053	11	AC002429	AC002429 Homo sapi
9	238.6	16.4	142178	35	AC007160	AC007160 Homo sapi
10	236	16.2	179757	42	AC006960	AC006960 Homo sapi
11	235	16.1	134578	34	AC004824	AC004824 Homo sapi
12	235	16.1	220000	35	AC005674	AC005674 Homo sapi
13	234.6	16.1	102200	9	HS30M3	AL031775 Human DNA
14	234.6	16.1	80507	11	AC004024	AC004024 Homo sapi
15	234.6	16.1	188477	34	AC004971	AC004971 Homo sapi
16	234	16.0	240000	35	AC006393	AC006393 Homo sapi
17	233.6	16.0	96625	10	HSAC000118	AC000118 Human BAC
18	233.6	16.0	135448	34	AC004900	AC004900 Homo sapi
19	232.8	16.0	39760	11	AC005253	AC005253 Homo sapi
20	232.8	16.0	46213	11	AC005387	AC005387 Homo sapi
21	232.8	16.0	42030	42	AC007565	AC007565 Homo sapi
22	232.2	15.9	45987	11	AC000028	AC000028 Homo sapi
23	232.2	15.9	45976	11	HS081031	U81031 Homo sapi
24	232.2	15.9	169026	35	AC007371	AC007371 Homo sapi
25	231.8	15.9	234596	34	HS94M16	297201 Homo sapi
26	231.8	15.9	124956	34	HS193N13	AL078600 Homo sapi
27	231.4	15.9	100000	9	AP000046	AP000046 Homo sapi
28	231.4	15.9	100000	9	AP000114	AP000114 Homo sapi
29	231.4	15.9	100000	9	AP000190	AP000190 Homo sapi
30	231.4	15.9	47559	9	AP000303	AP000303 Homo sapi
31	231.4	15.9	43012	35	AC007850	AC007850 Homo sapi
32	231.2	15.8	134970	11	AC002504	AC002504 Homo sapi
33	231.2	15.8	127587	34	AC005959	AC005959 Homo sapi
34	230.8	15.8	208643	34	HS44N10	297197 Homo sapi
35	230.6	15.8	82359	42	AC004922	AC004922 Homo sapi
36	230.6	15.8	82359	42	AC004922	AC004922 Homo sapi
37	230.4	15.8	79376	9	HS454G6	298750 Human DNA s
38	230.2	15.8	114784	10	HS1189B24	AL030996 Human DNA
39	230.2	15.8	164028	42	AC004890	AC004890 Homo sapi
40	230	15.8	84170	9	HS130M4	275887 Human DNA s
41	230	15.8	20733	10	HS356B7	AL031714 Human DNA
42	230	15.8	202004	11	H0AF001549	AF001549 Human Chr
43	229.8	15.8	143747	9	HS37M22	282206 Human DNA s
44	229.8	15.8	177236	34	AC005143	AC005143 Homo sapi
45	229.6	15.7	163871	9	HS158E12	AL031584 Human DNA

ALIGNMENTS

RESULT 1

AF135023
LOCUS AF135023 4740 bp DNA
DEFINITION Homo sapiens Kallikrein-like protein 1 KUK-L1 gene, partial cds.
ACCESSION AF135023
NID G4589272
VERSION AF135023.1 GI:4589272
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 4740)
AUTHORS Yousef,G.M., Luo,L.Y. and Diamandis,E.P.
TITLE Identification of novel human Kallikrein-like genes on chromosome 19q13.3-q13.4
JOURNAL Biochem. Biophys. Res. Commun. (1999) In press
REFERENCE 2 (bases 1 to 4740)
AUTHORS Yousef,G.M., Luo,L.Y. and Diamandis,E.P.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-1999) Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada
FEATURES Location/Qualifiers

```

source
1. .4740
/organism="Homo sapiens"
/db xref="taxon:9606"
/chromosome="19"
/map="19q13.3-q13.4"
join(<2263. .2425,2847. .3097,3181. .3317,4588. .>4740)
/product="kallikrein-like protein 1 KLR-LI"
join(<2263. .2425,2847. .3097,3181. .3317,4588. .4740)
/codon_start=3
/product="kallikrein-like protein 1 KLR-LI"
/protein_id="MAD26424.1"
/db_xref="PIR:G4589273"
/db_xref="GI:4589273"
/translation="SLVSSGSCQIINGEDCSPHSOPMAALVWENELFCGSLVHPDM
VLSAHCFOHSYTIIGLSTLEADDEPSOMVSLSVHPENRPLANDMLIKTD
EVSSEDPTRISISAOCPSTAGNSCLVSWGLANRMTVYLCNVSVSEVSKL
YPLVHPSMFCAGGDDQKDSGSDSGPLICNGYLDGLVSFGKAPCGGVPGYTTN
LCKTEWIKRTQAS"
BASE COUNT      1106 a      1275 c      1308 g      1051 t
ORIGIN

Query Match      96.78; Score 1411.4; DB 42; Length 4740;
Best Local Similarity 99.68; Pred. No. 0;
Matches 1423; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

33  CAGAGCTCTACACACATGGGGCTGGGCTGACAGTCTTGAGGCGGCAAGAGCCAGG 92
2842 CAGAGCTCTACACACATGGGGCTGGGCTGACAGTCTTGAGGCGGCAAGAGCCAGG 2801
93  AGCCAGATGTGAGAGGCGACCTCTCCGTACAGGACCCAGACAGTACACAGACCTTGCTC 152
2902 AGCCAGATGTGAGAGGCGACCTCTCCGTACAGGACCCAGACAGTACACAGACCTTGCTC 2961
153  GCTAAGACCTCATGCTCATCAAGTTGAGAGATCCGTGTCCAGTGTGACACCATCCGG 212
2962 GCTAAGACCTCATGCTCATCAAGTTGAGAGATCCGTGTCCAGTGTGACACCATCCGG 3021
213  AGCATCAGCATTCCTTCGCGAGTGCCTTACCGGGGGAACCTTGCCCTGTTTGGGGTGG 272
3022 AGCATCAGCATTCCTTCGCGAGTGCCTTACCGGGGGAACCTTGCCCTGTTTGGGGTGG 3081
273  GGTCTGCTGGCGAAGCGGTGAGCTACAGGGTGTGTCTGTGCTTCCCTTTCAGAGAGTCTCT 332
3082 GGTCTGCTGGCGAAGCGGTGAGCTACAGGGTGTGTCTGTGCTTCCCTTTCAGAGAGTCTCT 3141
333  GCCCAGTCCGGGGGGGCGTACAGGCTGTGCTGCCAGCAGAAATGCTTACCGTGTGC 392
3142 GCCCAGTCCGGGGGGGCGTACAGGCTGTGCTGCCAGCAGAAATGCTTACCGTGTGC 3201
393  AGTGCATGAGCTGTGGTGTCTGTGANGAGTCTCANTAAAGCTTATGACCCGCTGT 452
3202 AGTGCATGAGCTGTGGTGTCTGTGANGAGTCTCANTAAAGCTTATGACCCGCTGT 3261
453  ACCACCCCANCATGTTCTGTGCGCGCGGCGAGGCAAGCAGAAAGCATCTCTGCAAGCTGA 512
3262 ACCACCCCANCATGTTCTGTGCGCGCGGCGAGGCAAGCAGAAAGCATCTCTGCAAGCTGA 3321
513  GAGAGGGGAAAGGGAGGCGAGCTCAGGGAAGGGTGAAGAAAGGGGAGACAGAGAC 572
3322 GAGAGGGGAAAGGGAGGCGAGCTCAGGGAAGGGTGAAGAAAGGGGAGACAGAGAC 3381
573  ACACAGGGCGCATGGGAGATGCAGAGATGAGAGACACACAGGGGAGAGTGAAGTACT 632
3382 ACACAGGGCGCATGGGAGATGCAGAGATGAGAGACACACAGGGGAGAGTGAAGTACT 3441
633  AGAGAGAGAAACTGAGAGAAACAGAGAAATTAACACAGAAATTAAGAGAAAGCAAGGAG 692
3442 AGAGAGAGAAACTGAGAGAAACAGAGAAATTAACACAGAAATTAAGAGAAAGCAAGGAG 3501
693  AGAGAGAGAAACAGACATGGGAGGCGAGAAACACACACACATAGAAATGCAAGTTGACC 752
3502 AGAGAGAGAAACAGACATGGGAGGCGAGAAACACACACACATAGAAATGCAAGTTGACC 3561

```

```

753  TTCACACAGCATGGGGCGCTGAGGGGGGTGAGCTCCACCCCAATAGAAATCTCTATAC 812
3562 TTCACACACAGCATGGGGCGCTGAGGGGGGTGAGCTCCACCCCAATAGAAATCTCTATAC 3621
813  TTTTGAATCCCCCAAAACCTGACTAGAAAATAGCTACTGTTGACGGGAGGACCTTACCAAT 872
3622 TTTTGAATCCCCCAAAACCTGACTAGAAAATAGCTACTGTTGACGGGAGGACCTTACCAAT 3681
873  AACATAAATAGTCGATTTTATGATACGTTTATGATTCATATATACCTTTGTGGAAAT 932
3682 AACATAAATAGTCGATTTTATGATACGTTTATGATTCATATATACCTTTGTGGAAAT 3741
933  TTTTATATTTCTAGACACACAGTGTCTGTAATTTTAAATGTTGCAACTCT 992
3742 TTTTATATTTCTAGACACACAGTGTCTGTAATTTTAAATGTTGCAACTCT 3801
993  CCTAAATTTTCTGATGTGTTTATTTGAAAAAATCCAAATAGTATAGTGAAGTGTGCA-TT 1051
3802 CCTAAATTTTCTGATGTGTTTATTTGAAAAAATCCAAATAGTATAGTGAAGTGTGCAAGTT 3861
1052 CAACACAGGCTGTTTCAAGGCTCACTGTGTACCCAGAGGAAACAGATGACACAGATTCA 1111
3862 CAACACAGGCTGTTTCAAGGCTCACTGTGTACCCAGAGGAAACAGATGACACAGATTCA 3921
1112 TAGAGTGAACACAGAGAAAGAAACAGAAATCAAGACTCTACAAAGAGGCTGGGCGAG 1171
3922 TAGAGTGAACACAGAGAAAGAAACAGAAATCAAGACTCTACAAAGAGGCTGGGCGAG 3981
1172 GTGGCTCATGCTGTATATCCACAGACTTTGGAGGCGAGGAGAGATCACTTGNAGTA 1231
3982 GTGGCTCATGCTGTATATCCACAGACTTTGGAGGCGAGGAGAGATCACTTGNAGTA 4041
1232 AGGAGTTCAAGACACAGCTTGGCCAAATGCTGTGTGTACTTAAATATCAAAA 1291
4042 AGGAGTTCAAGACACAGCTTGGCCAAATGCTGTGTGTACTTAAATATCAAAA 4101
1292 GTTACGTGATATGTTGGAGGCGCTGTATCCAGCTACTTGGAGGCTAGGAGAGA 1351
4102 GTTACGTGATATGTTGGAGGCGCTGTATCCAGCTACTTGGAGGCTAGGAGAGA 4161
1352 GAATTCCTGAATATGAGGAGGAGAGTGAAGTGAATGATCAGACACTATATCTCC 1411
4162 GAATTCCTGAATATGAGGAGGAGAGTGAAGTGAATGATCAGACACTATATCTCC 4221
1412 AGCTGGGCAACAGAGTAAAGTCTGTCTCAAAAAAATCAAAAAA 1459
4222 AGCTGGGCAACAGAGTAAAGTCTGTCTCAAAAAAATCAAAAAA 4269

RESULT 2
AF148532      4385 bp      DNA      PRI      09-JUN-1999
LOCUS      AF148532
DEFINITION      Homo sapiens kallikrein 4 (KLR4) gene, complete cds.
ACCESSION      AF148532
NID      95020095
VERSION      AF148532.1 GI:5020095
KEYWORDS
SOURCE
ORGANISM      human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 4385)
Stephenson,S.A., Verity,K., Ashworth,L. and Clements,J.A.
Localization of a new prostate specific antigen-related serine
protease gene, KLR4, is evidence for an expanded human kallikrein
(KLR) gene family cluster on chromosome 19q13.3-13.4
JOURNAL
Unpublished
2 (bases 1 to 4385)
Stephenson,S.A., Ashworth,L. and Clements,J.A.
Direct Submission
Submitted (03-MAY-1999) School of Life Science, Queensland
University of Technology, George Street, Brisbane, Queensland 4001,

```


TITLE Molecular cloning and characterization of prostate, an androgen-regulated serine protease with prostate-restricted expression
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)
MEDLINE 99179024
REFERENCE 2 (bases 1 to 5900)
AUTHORS Nelson, P.S., Gan, L., Ferguson, C., Moss, P., Gellinas, R., Hood, L. and Wang, K.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th St SE, Bothell, WA 98021, USA
FEATURES
Source Location/Qualifiers
1..5900
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13"
join(<1411..1471,2735..2897,3319..3569,3653..3789,5062..5817)
/product="serine protease prostate"
1411..1471
/number=1
join(1411..1471,2735..2897,3319..3569,3653..3789,5062..5214)
/note="androgen regulated"
/codon_start=1
/product="serine protease prostate"
/protein_id="AAD21581.1"
/db_xref="PDB:4512032"
/db_xref="GI:4512032"
/translation="MATAGNPWFGLYLILIGVAGSLVSGSCQIINEEDSPHSQPM QALVMEMLFCSGVLYHPQWVLSAHCPSNTYIGLHSEADQEPSSQVYELAS VHPYRNRLIDMLIKLDESVESDTRISISASOCPITGNSCLVSGWGLANGR MPTLVQVNVSVSEVCSKLYDPLVHPMPAGGHDCKDSCNDSGGLPCNGYLO GLVSFGKACPGGVPGVYTNLCKFTMEIKTVQAS"
exon 2735..2897
exon 3319..3569
exon 3653..3789
exon 5062..5817
BASE COUNT 1417 a 1643 c 1569 g 1271 t
ORIGIN
Query Match 96.3%; Score 1405; DB 42; Length 5900;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1419; Conservative 0; Mismatches 8; Indels 1; Gaps 1;
33 CANAGCTCCACACCATGGGGCTGGCGTGCACAGTCTTGAGGCCGACCAAGGCCAGGG 92
|||
3314 CACAGTCTCCACACCATGGGGCTGGCGTGCACAGTCTTGAGGCCGACCAAGGCCAGGG 3373
QY 93 AGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCACAGATACAAACAGACCTTGCTC 152
|||
3374 AGCCAAATGGTGAAGGCCAGCCTCTCCGTACGGCACCACAGATACAAACAGACCTTGCTC 3433
DB 153 GCTAACGACCTCATGCTCAATGAATTGGAGCAATCCGTGTCCGAGTCTGACACCATCCGG 212
|||
3434 GCTTAAGACCTCATGCTCAATGAATTGGAGCAATCCGTGTCCGAGTCTGACACCATCCGG 3493
DB 213 AGCATAGCATTTGCTTCGACGTGCCCTACCGGGGGAACCTTGCTCGTTTGTGGCTGG 272
|||
3494 AGCATAGCATTTGCTTCGACGTGCCCTACCGGGGGAACCTTGCTCGTTTGTGGCTGG 3553
QY 273 GGTCTGCTGGCAAGCGTACGTCACGGGTGTGTCTGGCCTCTTCAAGAGAGTCTCT 332
|||
3554 GGTCTGCTGGCAAGCGTACGTCACGGGTGTGTCTGGCCTCTTCAAGAGAGTCTCT 3613
DB 333 GCCCACTCGCGGGGCTGACCCAGAGCTGTGCTCCAGGACAGAAATGCCCTACCGTCTGC 392
|||

DB 3614 GCCCACTCGCGGGGCTGACCCAGAGCTCTGCTGCCAGGACAGAAATGCCCTACCGTCTGC 3673
QY 393 AGTGGGTGAACGTGTCGGGTGTCGTGANGAGTCGTGANTAACTATACACCCGCTG 452
|||
3674 AGTGGGTGAACGTGTCGGGTGTCGTGANGAGTCGTGANTAACTATACACCCGCTG 3733
QY 453 ACCACCCCMCATGTTCTGCGCGCGGAGGAGGACAGACAGACGACTCTGCACGTGA 512
|||
3734 ACCACCCCMCATGTTCTGCGCGCGGAGGAGGACAGACAGACGACTCTGCACGTGA 3793
DB 513 GAGAGGGCAAGGGGAGGCGACGACTCAGGGAAGGTTGGAAGGGGAGACAGAGAC 572
|||
3794 GAGAGGGGCAAGGGGAGGCGACGACTCAGGGAAGGTTGGAAGGGGAGACAGAGAC 3853
QY 573 ACACAGGGCGCGATGGCGGATGACAGAGTGGAGACACACAGGAGACAGAGTGCACACT 632
|||
3854 ACACAGGGCGCGATGGCGGATGACAGAGTGGAGACACACAGGAGACAGAGTGCACACT 3913
DB 633 AGAGAGAGAACTGAGAGAAACAGAGAAATAAACAGAGAAATAAAGAGCAAGAGAG 692
|||
3914 AGAGAGAGAACTGAGAGAAACAGAGAAATAAACAGAGAAATAAAGAGCAAGAGAG 3973
QY 693 AGAGAAACAGAAACAGACATGGGGAGGCGAAGAACACACACATAGAAATGCAGTTGACC 752
|||
3974 AGAGAAACAGAAACAGACATGGGGAGGCGAAGAACACACACATAGAAATGCAGTTGACC 4033
QY 753 TTCCAAACACATGGGGCGCTGAGGGGGGTGACCTCCACCAATAGAAATCCCTTATAC 812
|||
4034 TTCCAAACACATGGGGCGCTGAGGGGGGTGACCTCCACCAATAGAAATCCCTTATAC 4093
DB 813 TTTTGACTCCCAAAACCTGACTAGAAATAGCTACTGTTGACGGGAGGCTTACCAAT 872
|||
4094 TTTTGACTCCCAAAACCTGACTAGAAATAGCTACTGTTGACGGGAGGCTTACCAAT 4153
QY 873 AACATTAATAGTCGATTTATGTCATACGTTTATGCAATTCATGATATACCTTTGTTGAAT 932
|||
4154 AACATTAATAGTCGATTTATGTCATACGTTTATGCAATTCATGATATACCTTTGTTGAAT 4213
QY 933 TTTTGATATTTCTAAGCTACACAGTGGTGTGAATTTTAAATGTTGCAACT 992
|||
4214 TTTTGATATTTCTAAGCTACACAGTGGTGTGAATTTTAAATGTTGCAACT 4273
DB 993 CCTAAATTTTCTGATGTTGTTATGAAAAATCAAGTATAGTGAAGTGTGCA--TT 1051
|||
4274 CCTAAATTTTCTGATGTTGTTATGAAAAATCAAGTATAGTGAAGTGTGCAAGT 4333
QY 1052 CAACACAGGTTGTTCAAGGTTCAACTGTGTACCCAGAGGGAACAGTGCACAGATTCA 1111
|||
4334 CAACACAGGTTGTTCAAGGTTCAACTGTGTACCCAGAGGGAACAGTGCACAGATTCA 4393
QY 1112 TTAGAGTGAACACGAAAGGAAGAAATCAAGACTGTACAAAGGCTGGGACGG 1171
|||
4394 TTAGAGTGAACACGAAAGGAAGAAATCAAGACTGTACAAAGGCTGGGACGG 4453
DB 1172 GTGGCTCATGCTGTATATCCAGCACTTGGGAGGCGAGGCGAGATCACTATAGATA 1231
|||
4454 GTGGCTCATGCTGTATATCCAGCACTTGGGAGGCGAGGCGAGATCACTATAGATA 4513
QY 1232 AGAGTTCAAGACACAGCTGGCCAAATGTTGAATCCTGTCTGACTTAAATAACAATA 1291
|||
4514 AGAGTTCAAGACACAGCTGGCCAAATGTTGAATCCTGTCTGACTTAAATAACAATA 4573
DB 1292 GTTAGCTGATATGTTGAGGCGGCTGTAAATCCAGCACTTGGGAGGCTGAGGACAGA 1351
|||
4574 GTTAGCTGATATGTTGAGGCGGCTGTAAATCCAGCACTTGGGAGGCTGAGGACAGA 4633
QY 1352 GAATTCCTGAATATGAGGCGAGAGTTGAAGTGAATGATACACACTATATCTCC 1411
|||
4634 GAATTCCTGAATATGAGGCGAGAGTTGAAGTGAATGATACACACTATATCTCC 4693
DB 1412 AGCTGGGCAACAGATAGACTGTCTCAAAAAAAAAAAAAAAAAAAAA 1459
|||
4694 AGCTGGGCAACAGATAGACTGTCTCAAAAAAAAAAAAAAAAAAAAA 4741


```
RESULT 4
AF113140 1347 bp mRNA PRI 25-MAR-1999
LOCUS Homo sapiens serine protease prostate mRNA, complete cds.
DEFINITION AF113140
ACCESSION 94512029
VERSION AF113140.1 GI:4512029
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1347)
AUTHORS Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and
Wang,K.
Molecular cloning and characterization of prostate, an
androgen-regulated serine protease with prostate-restricted
expression
Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)
JOURNAL 99179024
MEDLINE 2 (bases 1 to 1347)
REFERENCE Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and
Wang,K.
Direct Submission
Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th. St SE,
JOURNAL Bothell, WA 98021, USA
FEATURES
source
1. .1347
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13"
1. .765
/note="androgen regulated"
/codon_start=1
/product="serine protease prostate"
/protein_id="A021580.1"
/db_xref="PIR:G4512030"
/db_xref="GI:4512030"
/translat="MATAGNPMWFLGILLYAGSLVSGSQIINGEDCSPHSQPM
OALVANELFCGVLHPQWVLSAHCFOVSYTIGLHSLADEPSQVWEASLS
VRHPEYRPLANDMLIKLDESSESSTISASOCTAGNSCLVSGMGLANGR
MPVLOCVNVSVSEVSCSKLYDPLVHPSMFCAGSGHDQKDSNGSGPLICNGTLO
GLVSEKAPCGOVGVPGVYTNLCKFTWIEKTVQAS"
1327. .1332
POLYA_site 269 a 489 c 334 g 255 t
BASE COUNT
ORIGIN
Query Match 20.3%; Score 296.4; DB 42; Length 1347;
Best Local Similarity 81.0%; Pred. No. 2e-59;
Matches 393; Conservative 0; Mismatches 9; Indels 83; Gaps 1;
25 AGTAGTGCAGAGCTCTTACACCATGGGGCTGCACAGCTTGAGGCGGACCAAG 84
212 ACCTTTCCAGAACTCTTACACCATGGGGCTGCACAGCTTGAGGCGGACCAAG 271
85 AGCCAGGAGCCAGATGTTGAGAGCCAGCTTCCGTACGCGACCCAGAGTACAAGAC 144
272 AGCCAGGAGCCAGATGTTGAGAGCCAGCTTCCGTACGCGACCCAGAGTACAAGAC 331
145 CTTTGCCTGCTAACGACCTATGCTCATCAAGTTGACGAATCCGTCCGAGTCTACA 204
332 CTTTGCCTGCTAACGACCTATGCTCATCAAGTTGACGAATCCGTCCGAGTCTACA 391
205 CCATCCGAGCATCAGATGCTTCCGAGAGCCCTACCGGGGGGAACTCTTGCTCTTT 264
392 CCATCCGAGCATCAGATGCTTCCGAGAGCCCTACCGGGGGGAACTCTTGCTCTTT 451
265 CTGGCTGGGGTCTGCTGGCGAAGCGTACAGGGGTGTGTCTCTCCCTCTTCAAGGA 324
```

```
Db 452 CTGGCTGGGGTCTGCTGGCGAAC----- 475
QY 325 GGTCTCTGTGCCAGTGGCGGGGCTGACCCAGAGCTGTGGTCCAGGAGATGCTTAC 384
Db 475 -----GGCAGATGCTTAC 488
QY 385 CGTCTGCAGTGCCTGTAACGTGTGGTGTCTGTGANGAGTCTGCANTAACTATGA 444
Db 489 CGTCTGCAGTGCCTGTAACGTGTGGTGTCTGTGANGAGTGTGAGTAACTATGA 548
QY 445 CCCGCTGTACACCCCAACATGTTTCCGCCGGCGGAGGCGAAGACAGAGACTCTTG 504
Db 549 CCCGCTGTACACCCCAACATGTTTCCGCCGGCGGAGGCGAAGACAGAGACTCTTG 608
QY 505 CAACG 509
Db 609 CAACG 613
RESULT 5
LOCUS HS310013 166723 bp DNA HTG 10-JUN-1999
DEFINITION Homo sapiens chromosome 20 clone 310013, WORKING DRAFT SEQUENCE, in
unordered pieces.
ACCESSION AL031658
NID 94902630
VERSION AL031658.10 GI:4902630
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 166723)
AUTHORS Kay,M.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On May 27, 1999 this sequence version replaced gi:4741461.
COMMENT IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Unfinished: dj310013 Contig_ID: 00232 acc-AL031658
Length: 1113 bp Unfinished: dj310013 Contig_ID: 00912
acc-AL031658 Length: 162976 bp Unfinished: dj310013 Contig_ID:
02100 acc-AL031658 Length: 1034 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
source
1. .166723
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="310013"
BASE COUNT 42104 a 40443 c 39486 g 43089 t 1601 others
ORIGIN
Query Match 17.0%; Score 248.2; DB 34; Length 166723;
Best Local Similarity 84.3%; Pred. No. 4.8e-48;
Matches 291; Conservative 0; Mismatches 53; Indels 1; Gaps 1;
1113 AGAGGTGAACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1172
Db 93064 AATTTTCAAACTCTTAAAGATAGACTTCAAAACATTCGATGTTGGGATATG 93123
QY 1173 TGCGTATCCCTTAATCCAGCATTTGGAGG-CGAGGACGAGCATCTTGAAGTA 1231
```

Db 93124 TGGCTCATGCTGTATCCAGCACTTTGGAGCCGAGGTGGGAAGATCACTGAGTTC 93183
 QY 1222 AGGAGTTCAAGACCGCCGCGCAAAATGGTAAATCCGTGTGTAATAAATCAAA 1291
 Db 93184 AGGAGTTCAAGACCGCCGCGCAAAATGGTAAATCCGTGTGTAATAAATCAAA 93243
 QY 1292 GTTACCTGATATGTTGGAGCGCGCTGTATCCAGCTACTTGGAGGCTGAGCAGGA 1351
 Db 93244 ATTACTGGGCGATGGTGGCAGACCGCTGTAGTCCAGCTACACAGAGGCTGAGCAGGA 93303
 QY 1352 GAATTGCTTGAATATGGAGCGAGAGGTTGAAGTGAAGTGAATCAACCACTATCTCC 1411
 Db 93304 GAATTGCTTGAACCTGGAGGCGGAGGTGCAAGTGAAGTCAACCACTATCTCC 93363
 QY 1412 ACTGGGCGACAGAGTAAGTGTCTGTCACAAAATAAAAAA 1456
 Db 93364 ACCCTGGCGCAGAGTGAAGTGTCTGTCACAAAATAAAAAA 93408

RESULT 6
 LOCUS HSAC002066 151770 bp DNA PRI 09-MAY-1997
 DEFINITION Human BAC clone RG030H15 from 7q31, complete sequence.
 ACCESSION AC002066
 NID 92076721
 VERSION AC002066.1 GI:2076721
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 151770)
 TITLE Dante, M., Mamsley, P. and Keppler, D.
 JOURNAL The sequence of H. sapiens BAC clone RG030H15
 REFERENCE 2 (bases 1 to 151770)
 TITLE Waterston, R.
 JOURNAL Direct Submisison
 COMMENT Submitted (09-MAY-1997)
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63108, USA
 http://genome.wustl.edu/gsc
 e-mail: sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7> or send an E-mail to egreen@hprt1.nih.gov

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).
 VECTOR: pBEO
 Selection: Chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
 Actual start of this clone is at base position 1 of H_RG030H15;
 actual end is at 151770 of H_RG030H15. The orientation of this clone is unknown.

This clone contains STS SWSS965.

FEATURES	Location/Qualifiers
source	1..151770
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/clone="RG030H15"
	/clone_1id="CTB-978SK-B"
	/map="7q31"
repeat_region	complement(671..690)
	/rpt_family="L1"
repeat_region	2544..2593
	/rpt_family="L1"
repeat_region	complement(2759..2781)
	/rpt_family="L1"
repeat_region	complement(4347..4468)
	/rpt_family="L1"
repeat_region	complement(5317..5615)
	/rpt_family="ALU"
repeat_region	complement(6063..6314)
	/rpt_family="ALU"
repeat_region	complement(7577..8152)
	/rpt_family="L1"
repeat_region	complement(8176..8233)
	/rpt_family="L1"
repeat_region	complement(8706..9002)
	/rpt_family="ALU"
repeat_region	complement(9318..9390)
	/rpt_family="ALU"
repeat_region	complement(9403..9434)
	/rpt_family="L1"
repeat_region	complement(9481..9510)
	/rpt_family="L1"
repeat_region	9513..9789
	/rpt_family="ALU"
repeat_region	10325..10770
	/rpt_family="L1"
misc_feature	12726..13009
	/note="similar to human EST D45516 (NID:9662470)"
repeat_region	14161..15579
	/rpt_family="L1"
repeat_region	15804..15902
	/rpt_family="L1"
repeat_region	17101..17218
	/rpt_family="ALU"
repeat_region	complement(17462..17485)
	/rpt_family="L1"
repeat_region	complement(17519..17675)
	/rpt_family="ALU"
repeat_region	complement(18866..19055)
	/rpt_family="L1"
repeat_region	complement(19068..20712)
	/rpt_family="L1"
repeat_region	complement(20933..21001)
	/rpt_family="L1"
repeat_region	complement(21021..21785)
	/rpt_family="L1"
repeat_region	complement(21919..21943)
	/rpt_family="L1"
repeat_region	23706..24283
	/rpt_family="L1"
repeat_region	25651..25942
	/rpt_family="ALU"
repeat_region	27166..27287
	/rpt_family="ALU"
repeat_region	27292..27794
	/rpt_family="L1"
repeat_region	complement(27974..28348)

repeat_region	/rpt_family="I1"	28315 .28348
repeat_region	/rpt_family="I1"	complement(30214 .30233)
repeat_region	/rpt_family="I1"	complement(30338 .30564)
repeat_region	/rpt_family="I1"	30584 .30882
repeat_region	/rpt_family="ALU"	30894 .30920
repeat_region	/rpt_family="I1"	complement(30906 .31081)
repeat_region	/rpt_family="I1"	complement(31082 .31447)
repeat_region	/rpt_family="ALU"	complement(31833 .32074)
repeat_region	/rpt_family="I1"	complement(32115 .32165)
repeat_region	/rpt_family="I1"	complement(32333 .33181)
misc_feature	/rpt_family="I1"	34263 .36923
misc_feature	/note="Diffuse retroviral homology. See Z76735 (NID:g1438501)." complement(34263 .34413)	
misc_feature	/note="similar to human EST AA151971 (NID:g1720827 zo30b03.r1"	
misc_feature	complement(34263 .34489)	
misc_feature	/note="similar to human EST C18895 (NID:g1580497)" complement(34263 .34526)	
misc_feature	/note="similar to human EST AA133084 (NID:g1694609 zo22h03.r1"	
misc_feature	complement(34474 .34547)	
misc_feature	/note="similar to human EST AA132265 (NID:g1693753 zo27all.r1"	
misc_feature	complement(34549 .34671)	
misc_feature	/note="similar to human EST AA055687 (NID:g1548225 z1l6h06.r1"	
misc_feature	complement(34549 .34700)	
misc_feature	/note="similar to human EST AA132265 (NID:g1693753 zo27all.r1"	
misc_feature	complement(34549 .34666)	
misc_feature	/note="similar to human EST AA134504 (NID:g1692069 zo12a05.r1"	
misc_feature	complement(34659 .34775)	
misc_feature	/note="similar to human EST AA152341 (NID:g1719175 zo01g11.r1"	
misc_feature	complement(34762 .35004)	
misc_feature	/note="similar to human EST C18935 (NID:g1580537)" complement(34881 .34991)	
misc_feature	/note="similar to human EST C04699 (NID:g1467950)" complement(35024 .35255)	
misc_feature	/note="similar to human EST AA099408 (NID:g1645301 z1l7a05.s1"	
misc_feature	complement(35025 .35255)	
misc_feature	/note="similar to human EST T51824 (NID:g653684) yb54b04.s1"	
misc_feature	35321 .35361	
misc_feature	/note="similar to human EST AA160337 (NID:g1734916 zo56e05.s1"	
misc_feature	35361 .35669	
misc_feature	/note="similar to human EST T03007 (NID:g457094)" complement(35694 .36004)	
misc_feature	/note="similar to human EST AA135325 (NID:g1697193 zo28e05.r1"	
misc_feature	complement(36078 .36282)	
misc_feature	/note="similar to human EST AA152341 (NID:g1719175 zo01g11.r1"	
misc_feature	complement(36078 .36414)	
misc_feature	/note="similar to human EST AA149780 (NID:g1720723 zo01a08.r1"	
misc_feature	36141 .36342	
misc_feature	/note="similar to human EST T61055 (NID:g664092)	

	misc_feature	yb75e05.s1*
		complement(36155..36653)
Query Match	16.5%	Score 240.6; DB 10; Length 151770;
Best Local Similarity	55.7%;	Pred. No. 2.8e-46;
Matches 504; Conservative	0;	Mismatches 394; Indels 7; Gaps 2;
OY	552	GAGACAGACACACAGGCGCCGATGGCAGATGTGCAGAGATGGAGAGACACACAGGGAGA 621
Db	64499	GAGAGAGAGAGAGGGAACAGAGAGAGAGAGTGAAGAGAAAAAGAGAGAGACAGGGAGA 64558
OY	622	CAGTCACAACCTAGAGAGAAACCTGAGAAAACAGAGAAATTAAACACAGGATTAAGAGA 681
Db	64559	GAGAGAGAGAGAAAGAGAGAGAAAAACAGAGAGAGAGAGAGTGAAGAGAAAAAGAGAGA 64618
OY	682	AGCAAGAGNAGAGAAAAACAGAACAGACATGTGGGAGCAGAAACACACACATAGANA 741
Db	64619	GAGAGAGACAGGAGAGAGAGAGAGAGAGAAAAGAGAGAAAAACAGAGAGAGAGAGAG 64678
OY	742	TGCAGTTGACCTTCCACAGCATGGGGCGGTGAGGGGTGACTCCACCATTAGAAAAT 801
Db	64679	GAGAGAGAACAGAGAGATGCAGAGACATCATAGATAGAGAGGTGCCAGAGCAGAGAA 64738
OY	802	CCTCTATTAACCTTTTGATCCCCCAAACCTGACTAGAAATAGCCTACTGTTGAGGGGA 861
Db	64739	CCCTCAACCCCTAAAAAGAAAGTCACTAACAGTCTGCTGTCAGTTGGCCGTGATGAT 64798
OY	862	GCCTTAACAATTAATAATAGTCATTTATGATACAGTTTATGTCATTATCATGATATACC 921
Db	64799	AGGCCACCCCTTCTCCATTTATTAACGATGACAAACCTCTATATATAAAGCCGTAAACC 64858
OY	922	TTTGTGGAATTTTTTGATATTTCTAAGCTCACAGTTGCTGTGAAATTTTTTAAAT 981
Db	64859	AGCATTTGCAAAATAGAGATTTCTCAAAACATACAGAGATACCCAACACTGAGAAAAC 64918
OY	982	GTTGCACTCTCTTAAATTTTTTCGATGTGTTATTGAAAAATCCAGATTTAAGTGA 1041
Db	64919	AAATGAGAGAAATAGAACTGATAGAAACAGAAATCTTAAAAAATATATATTATAT 64978
OY	1042	CTTGCAATTCAAACAGGGGTGTTCAAGGGGCACTGTGTCACCAAGAGGAAACAGTGA 1101
Db	64979	CCCTAAAAACACTCAAGAGATATGCAATTAATTAATTAAGACCAAGAAATTTCTAAAGA 65038
OY	1102	CACAGATTCATAGAGGTGAACACGAGAGAGAAACAGAAA-----AATCAAGACTTAC 1155
Db	65039	AAATATATGTCAAAAAAGTAAACATGTATGAGGTCAAAATACTTATCAATCAAGATTA 65098
OY	1156	AAAGAGCTGGGCGAGGGTGGCTCATGCTGTATATCCAGCACTTTGGAGGC-GAGGCAG 1214
Db	65099	AAATTAAGCTGGGTGGGGGTGCTCAAGCCCTGTATCCAGCACTTTGGGGGCTGAGGCAG 65158
OY	1215	GCAGATCACTTAGGTAAAGAGTTCAANACAGCCCTGGCCAAATGGTAAATCCTGTCT 1274
Db	65159	GGAGATCACTAGGTCTCAAGGTTTGAAGAGCAGCCTGGCCAACATGTGAACCCCACT 65218
OY	1275	GTACTAAAAATACAAAAGTTAGCTGATATGGTGGCAGGCGCCTTAATCCAGCTACTT 1334
Db	65219	CTACTAAAAATACAAAATTTAGCCAGGTGTGGGGTGGGCGCTGTATCCAGCTATACC 65278
OY	1335	GGGAGGCTGGAGCGAGCAATTGCTTGAATATGGAGCGACAGGTTGAAGTGAAGTTGAGA 1394
Db	65279	GGGAGGCTGGAGGCTGGAATAATCACTTGAACCCGGGAGGACAGAGGTTGCAAGTGA 65338
OY	1395	TCACACCACTTAATCCACAGCTGGGGCACAAGAGTAAGACTGCTCAAAAAAAAAAAAA 1454
Db	65339	TCACACCACTGCACTCCAGCCTGGGTGAACAGCTAGACTCCGTCTCAAAAAAAAAAAAA 65398
OY	1455	AAAAA 1459
Db	65399	AAAAA 65403
RESULT		7

	DB	111608	GCGTGTATCCACAGACTTTGGGAGGCTGACGCAGCAGCAGACATCACTTAGTGTCAGAGATT	111549
QY	1240	AAGACCAGCTTGCGCAAATGCTGAATCTGTCTGTACTATAAAATACAAAATTAGCTG	1299	
Db	111548	GAGACACGCGCTGTCATATGTTGTAACCTGTCTACTAAAAATACAAAATTAGCTG	111489	
Qy	1300	GATATGGTGGCAGGCGCGCTGTAACTCCAGCTACTTGGGAGGCTGAGCAGGAAATGCT	1359	
Db	111488	GGTGTGGTGGCAGCAGCTGTAACTCCAGCTACTTGGGAGGCTGAGCAGGAAATGCT	111429	
QY	1360	TGAATATGGGAGGAGCAGAGTTGAAGTTGAGATCATCACACACTATACCTCAGCTGGG	1419	
Db	111428	TGAACCTGGGAGGCGGAGGTTGCAGTGAAGCTGACATGCACACATTCGACCTCAGCTRAG	111369	
QY	1420	CACAGAGCTAAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAA	1459	
Db	111368	CACAGAGCTAAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAA	111329	
RESULT	8			
AC002429				
LOCUS	AC002429	234053 bp	DNA	PRI 19-AUG-1997
DEFINITION	Human BAC clone GS200K05 from 7q21-q22, complete sequence.			
ACCESSION	AC002429			
NID	G2335067			
VERSION	AC002429.1	GI:2335067		
KEYWORDS	Htg.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Granata; Vertebrata; Mammalia;			
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 234053)			
JOURNAL	Graves,T, Duckels,G and Hawkins,M.			
REFERENCE	The sequence of H. sapiens BAC clone GS200K05			
AUTHORS	Unpublished (1997)			
TITLE	2 (bases 1 to 234053)			
JOURNAL	Waterston,R.			
COMMENT	Direct Submission Submitted (19-AUG-1997) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA SUBMITTED BY: Genome Sequencing Center Department of Genetics Washington University St. Louis MO 63108, USA http://genome.wustl.edu/gsc mailto:sapiens@watson.wustl.edu			

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GTB/CHRT> or send mailto:egreen@nhgri.nih.gov

SOURCE INFORMATION:
This clone is from the first BAC library from Genome Systems, Inc. (<http://www.genomesystems.com>).

Cell line: lymphoblastoid
Haplotypes: two
VECTOR: pBELOBAC
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 1 of 1 GS200K05;
actual end is at 234053 of GS200K05. The orientation of this clone
is unknown.

This cosmid contains sts's SWSS3537 (NID:g1222821) and SWSS3746
(NID:g1916314).

FEATURES

SOURCE	Location/Qualifiers
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="7"	
/clone="GS200K05"	
/map="7q21-q22"	
/map="7q21-q22"	
complement(47..74)	
/rpt_family="L1"	
complement(145..171)	
/rpt_family="L1"	
1042..1102	
/rpt_family="L1"	
complement(3234..3271)	
/rpt_family="L1"	
complement(3293..3329)	
/rpt_family="L1"	
complement(3345..3560)	
/rpt_family="L1"	
complement(4040..7521)	
/rpt_family="L1"	
5186..5603	
/rpt_family="L1"	
5764..5791	
/rpt_family="L1"	
complement(7522..7598)	
/rpt_family="ALU"	
complement(7599..8058)	
/rpt_family="L1"	
complement(8059..8335)	
/rpt_family="ALU"	
complement(8336..10510)	
/rpt_family="L1"	
complement(10522..10654)	
/rpt_family="ALU"	
10731..10962	
/note="match to EST AA011406 (NID:g1472432) z101901.s1"	
complement(11288..11342)	
/rpt_family="L1"	
complement(12291..12332)	
/rpt_family="L1"	
12823..12860	
/rpt_family="L1"	
complement(13091..13119)	
/rpt_family="L1"	
complement(13605..13894)	
/rpt_family="ALU"	
14412..14433	
/rpt_family="L1"	
complement(16228..16474)	
/rpt_family="ALU"	
17447..17474	
/rpt_family="L1"	
complement(17649..17938)	
/rpt_family="ALU"	
complement(18683..18974)	
/rpt_family="ALU"	
complement(20349..20389)	
/rpt_family="ALU"	
complement(20398..20501)	

repeat_region	/rpt_family="ALU"	complement(21120..21414)
repeat_region	/rpt_family="ALU"	complement(21722..25632)
repeat_region	/rpt_family="L1"	23274..23301
repeat_region	/rpt_family="L1"	26198..26636
repeat_region	/rpt_family="L1"	complement(27632..27653)
repeat_region	/rpt_family="L1"	complement(27776..27803)
repeat_region	/rpt_family="L1"	28127..28151
repeat_region	/rpt_family="L1"	29781..29826
repeat_region	/rpt_family="L1"	complement(30309..30326)
repeat_region	/rpt_family="L1"	complement(30344..31472)
repeat_region	/rpt_family="L1"	complement(31542..31829)
repeat_region	/rpt_family="ALU"	complement(32096..32133)
repeat_region	/rpt_family="L1"	complement(32165..32729)
repeat_region	/rpt_family="L1"	complement(32758..33562)
repeat_region	/rpt_family="L1"	complement(34455..34486)
repeat_region	/rpt_family="L1"	34492..34778
repeat_region	/rpt_family="ALU"	complement(35297..35386)
repeat_region	/rpt_family="L1"	35973..36329
repeat_region	/rpt_family="L1"	complement(36687..36981)
repeat_region	/rpt_family="ALU"	37162..37200
repeat_region	/rpt_family="L1"	complement(38651..38925)
repeat_region	/rpt_family="ALU"	complement(39644..39687)
repeat_region	/rpt_family="L1"	complement(39688..39980)
repeat_region	/rpt_family="ALU"	complement(40197..40226)
repeat_region	/rpt_family="L1"	41164..41184
repeat_region	/rpt_family="L1"	41716..41765
repeat_region	/rpt_family="L1"	43720..43753
repeat_region	/rpt_family="L1"	45223..45251
repeat_region	/rpt_family="L1"	complement(46066..46102)
repeat_region	/rpt_family="L1"	complement(46105..46388)
repeat_region	/rpt_family="ALU"	complement(46748..47039)
repeat_region	/rpt_family="ALU"	47669..47958
repeat_region	/rpt_family="ALU"	48675..49195
repeat_region	/rpt_family="L1"	49988..50277
repeat_region	/rpt_family="ALU"	complement(50408..50442)
repeat_region	/rpt_family="L1"	complement(53772..53879)
repeat_region	/rpt_family="L1"	

repeat_region complement(55270..55317)
/rpl_family="I1"

Query Match 16.4%: Score 238.8; DB 11; Length 234053;
Best Local Similarity 83.0%: Pred. No. 7.6e-46;
Matches 284; Conservative 0; Mismatches 57; Indels 1; Gaps 1;

QY 1119 GAAACAGAGAGAAACAGAGAAATCAAGACTCTACAAAGAGGCTGGCAGGTTGGCTC 1178
DB 134434 GAAGAGGTAGTAGAACAAGATCCTTTAAGATAGAAACAGGCTGGCAGTGTGCTC 134493
QY 1179 ATGCTGTATCCAGCAGCTTTGGAGG-CGAGGAGGAGATCCTTGAAGTAAAGT 1237
DB 134494 ACACCTGTATCCAGCAGCTTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 134553
QY 1238 TCAAGACGAGCTGGGCAAAATGGTGAATCCTGTCTACTATAAATCAAAAGTTAGC 1297
DB 134554 TTGAGACGAGCTGGGCAAAATGGTGAATCCTGTCTACTATAAATCAAAAGTTAGC 134613
QY 1298 TGGATGTGGCAGGAGGCTGTATCCAGCTACTTGGAGGCTGAGGAGGAGGAGGAGG 1357
DB 134614 CGGCGATGGTGGCAGGAGGCTGTATCCAGCTACTTGGAGGCTGAGGAGGAGGAGG 134673
QY 1358 CTTGAATATGAGAGGAGGTTGAAGTGAAGTGAATGATGATGATGATGATGATGATG 1417
DB 134674 CTTGAATATGAGAGGAGGTTGAAGTGAAGTGAATGATGATGATGATGATGATGATG 134733
QY 1418 GGCACAGAGTAAAGCTGTCTCAAAAAAATCAAAAAA 1459
DB 134734 GGCACAGAGGAGGAGGAGGCTGTCTCAAAAAAATCAAAAAA 134775

RESULT 9

LOCUS AC007160 142178 bp DNA HTG 07-APR-1999
DEFINITION Homo sapiens clone DJ0146103, WORKING DRAFT SEQUENCE, 2 unordered
ACCESSION AC007160
NID 94572709
VERSION AC007160.2 GI:4572709
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 142178)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 142178)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source
1..49134: contig of 49134 bp in length
* 49135 49152: gap of unknown length
* 49153 142178: contig of 93026 bp in length.
Location/Qualifiers
1..142178
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DJ0146103"

BASE COUNT 39502 a 33809 c 32673 g 36176 t 18 others

ORIGIN

Query Match 16.4%: Score 238.6; DB 35; Length 142178;
Best Local Similarity 84.8%: Pred. No. 8.1e-46;
Matches 279; Conservative 0; Mismatches 49; Indels 1; Gaps 1;

QY 1131 GAAACAGAGAAATCAAGACTCTACAAAGAGGCTGGCAGGTTGGCTC 1190
DB 31842 GAAATCAAAATGATAGAACTCTACAAAGAGGCTGGCAGGTTGGCTC 31901
QY 1191 CCAGCAGCTTTGGAGG-CGAGGAGGAGATCCTTGAAGTAAAGTAAAGTAAAGTAAAGT 1249
DB 31902 CCAGCAGCTTTGGAGGAGGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 31961
QY 1250 TGGCAAAATGGTGAATCCTGTCTACTATAAATCAAAAGTAAAGTAAAGTAAAGTAAAGT 1309
DB 31962 TGGCAAAATGGTGAATCCTGTCTACTATAAATCAAAAGTAAAGTAAAGTAAAGTAAAGT 32021
QY 1310 CAGGAGCTGTATCCAGCTACTTGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1369
DB 32022 CAGGAGCTGTATCCAGCTACTTGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 32081
QY 1370 AGCAGAGGTTGAAGTGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATG 1429
DB 32082 AGGAGAGGTTGAAGTGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATG 32141
QY 1430 AGACTGTCTCAAAAAAATCAAAAAA 1458
DB 32142 AGACTGTCTCAAAAAAATCAAAAAA 32170

RESULT 10

LOCUS AC006960 179757 bp DNA PRI 05-MAR-1999
DEFINITION Homo sapiens clone UWGC:djs58 from 7p14-15, complete sequence.
ACCESSION AC006960
NID 94337211
VERSION AC006960.1 GI:4337211
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 179757)
AUTHORS Iadonato,S.P., Yu,J., Wong,G.K.-S., Magness,C.L., Green,E.D.,
Green,P. and Olson,M.V.
TITLE Large-scale MCD Mapping and Sequencing of Human Chromosome 7
JOURNAL Unpublished (1999)
REFERENCE 2 (bases 1 to 179757)
AUTHORS Bubb,K.L. and Desmarais,C.L.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-1999) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, USA
University of Washington Human Genome Center
Box 352145 Seattle, WA 98195
Contact: Kerry Bubb (kbubb@u.washington.edu)

REMARK

COMMENT

Overlapping Sequences:
5': UWGC:djs77
3': UWGC:mapping in progress

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

Double stranded (DS) coverage: 99.7%
DS or two chemistry coverage: 99.9%

Single stranded regions: 1

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. Small fragments below a variable cutoff (approximately 400-600) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

BglII	Seq	Map	Seq	Map	Seq
1221.00	1225.00	777.00	753.00	2434.00	2433.00
9301.00	9214.00	2717.00	2724.00	2016.00	2016.00
2260.00	2247.00	6542.00	6545.00	958.00	925.00
542.00	540.00	2717.00	2671.00	15887.00	15655.00
2421.00	2383.00	3666.00	3619.00	1772.00	1786.00
4040.00	4054.00	2020.00	1958.00	2230.00	2233.00
1221.00	1208.00	3319.00	3354.00	5045.00	4995.00
1068.00	1047.00	7389.00	7370.00	527.00	518.00
1482.00	1487.00	847.00	842.00	1451.00	1421.00
9301.00	9202.00	3319.00	3252.00	958.00	945.00
5493.00	5469.00	10078.00	9931.00	5371.00	5299.00
4200.00	4199.00	5656.00	5592.00	1002.00	994.00
2816.00	2790.00	1896.00	1892.00	5045.00	4907.00
3863.00	3891.00	10078.00	10220.00	1072.00	1089.00
2616.00	2589.00	777.00	799.00	1002.00	999.00
7106.00	7198.00	6772.00	6793.00	3998.00	3967.00
8828.00	8842.00	2020.00	2027.00	1532.00	1507.00
5679.00	5807.00	2311.00	2325.00	2434.00	2427.00
1389.00	1385.00	11025.00	11100.00	1072.00	1080.00
5679.00	5685.00	8690.00	8656.00	2016.00	2029.00
13683.00	13752.00	2897.00	2881.00	3845.00	3855.00
7777.00	7675.00	736.00	714.00	3180.00	3192.00
519.00	508.00	14213.00	14297.00	1451.00	1448.00
769.00	759.00	1041.00	1082.00	5371.00	5459.00
33833.00	30657.00	5564.00	5453.00	646.00	632.00
5276.00	5201.00	1659.00	1648.00	748.00	737.00
2816.00	2771.00	26979.00	24815.00	4445.00	4419.00
7777.00	7750.00	6242.00	6209.00	2326.00	2308.00
2935.00	2913.00	1041.00	1024.00	5045.00	5023.00
2816.00	2827.00	3666.00	3630.00	8708.00	8619.00

1221.00	1210.00	3949.00	3923.00	527.00	531.00
890.00	878.00	777.00	767.00	791.00	791.00
3369.00	3317.00	1483.00	1470.00	646.00	643.00
658.00	641.00	3949.00	3851.00	3650.00	3525.00
19257.00		5656.00	5567.00	1532.00	1538.00
2071.00		930.00	917.00	4804.00	4766.00
2816.00		930.00	918.00	3650.00	3699.00
		4033.00	3969.00	5794.00	5678.00
		1041.00	1017.00	17463.00	17311.00
		8942.00		2690.00	2655.00
				1864.00	1835.00
				1532.00	1532.00
				1451.00	1437.00
				14855.00	14780.00
				6080.00	6005.00
				1532.00	1482.00
				28565.00	

Unmatched fragments are due to vector-containing sequences. We expect 3, 1 and 1-2 such fragments in the BglII, EcoRI, and NsiI enzyme domains, respectively. The sum of these fragments can be calculated from the sequence of the BAC clone and, in this clone, should be 26,152bp, 9,439bp, and 25,563bp for the three enzyme domains.

FEATURES Location/Qualifiers

source

1. 179757
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7p14-15"
 /clone="De Jong: dj182J23"
 /cell_line="Male blood"
 /clone_lib="De Jong BAC Library"
 2009..2065
 /rpt_family="Alu"
 3184..3272
 /rpt_family="MER20"
 3891..4165
 /rpt_family="Alu"
 complement(4937..5211)
 /rpt_family="Alu"
 complement(6618..6908)
 /rpt_family="Alu"
 complement(7575..7871)
 /rpt_family="Alu"
 complement(8397..8669)
 /rpt_family="Alu"
 10475..10554
 /rpt_family="MLT1"
 complement(13453..13492)
 /rpt_family="Alu"
 14584..14861
 /rpt_family="Alu"
 complement(15593..15880)
 /rpt_family="Alu"
 16455..16794
 /rpt_family="RHE1"

```

repeat_region 16799..18673
                /rpt_family="MSTAR"
repeat_region complement(17201..17498)
                /rpt_family="Alu"
repeat_region 17925..17966
                /rpt_family="THE1"
repeat_region 18711..19051
                /rpt_family="THE1"
repeat_region 19606..19875
                /rpt_family="Alu"
repeat_region 19885..20236
                /rpt_family="THE1"
repeat_region complement(23443..23690)
                /rpt_family="Alu"
repeat_region complement(23917..24133)

```

Query Match 16.2%; Score 236; DB 42; Length 179757;
 Best Local Similarity 84.5%; Pred. No. 3.3e-45;
 Matches 277; Conservative 0; Mismatches 50; Indels 1; Gaps 1;

```

QY 1132 AACAGAAATCAAGACTCTACAAAGAGCGTGGCGGCTCATGCCGTATCC 1191
    || || || || || || || || || || || || || || || || || || ||
DB 50074 AAATGATACCTTAAAGATTAACAGAGCGCTGGTATGTCACACCTGTATCC 50015
QY 1192 CAGCACTTTGGGAG-6CGAGGAGGAGATCACTTGAAGAGCTCAAGACGAGCT 1250
    || || || || || || || || || || || || || || || || || || ||
DB 50014 CAGCACTTTGGGAGTCCGAGGAGCGGAGATCACTTGAAGAGTCAAGAGCT 49955
QY 1251 GGGCAAAATGCGAATCTCTCTACTTAAATCAAAAGTACTGATATGCTGC 1310
    || || || || || || || || || || || || || || || || || || ||
DB 49954 GACCAACATGAGTAAACCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 49895
QY 1311 AGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1370
    || || || || || || || || || || || || || || || || || || ||
DB 49894 GGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 49835
QY 1371 GGCAGAGTTGAAGTGTGAGATCAACACCTATCTCTCTCTCTCTCTCTCTCTCT 1430
    || || || || || || || || || || || || || || || || || || ||
DB 49834 GGTGGAGGTGTACGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 49775
QY 1431 GACTGTCTCAAAAAAAAAAAAAAAAAAAAA 1458
    || || || || || || || || || || || || || || || || || || ||
DB 49774 GACACTGTCTCAAAAAAAAAAAAAAAAAAAAA 49747

```

RESULT 11
 AC004824 134578 bp DNA HTG 12-JUN-1998
 LOCUS AC004824
 DEFINITION Homo sapiens clone D020B21, WORKING DRAFT SEQUENCE, 4 unordered
 pieces.
 AC004824
 NID 93213173
 VERSION AC004824.1 GI:3213173
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 134578)
 Waterston,R.H.
 The sequence of Homo sapiens clone
 Unpublished
 2 (bases 1 to 134578)
 Waterston,R.H.
 Direct Submission
 Submitted (12-JUN-1998) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

COMMENT
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.

```

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 6892: contig of 6892 bp in length
* 6893 6909: gap of unknown length
* 6910 17964: contig of 11035 bp in length
* 17945 17961: gap of unknown length
* 17962 34519: contig of 16558 bp in length
* 34520 34536: gap of unknown length
* 34537 134578: contig of 100042 bp in length.
FEATURES
    source
        1..134578
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="D020B21"
BASE COUNT 32479 a 34383 c 33834 g 33831 t 51 others
ORIGIN

```

Query Match 16.1%; Score 235; DB 34; Length 134578;
 Best Local Similarity 84.4%; Pred. No. 5.5e-45;
 Matches 276; Conservative 0; Mismatches 50; Indels 1; Gaps 1;

```

QY 1134 ACAGAAATCAAGACTCTACAAAGAGCGTGGCGGCTCATGCCGTATCC 1193
    || || || || || || || || || || || || || || || || || || ||
DB 11958 ACATAAATTTAAATTAAGCTGGGAGCGGCTGAGCTCAGCTGTATCC 12017
QY 1194 GCACCTTTGGGAGG-6GAGGAGGAGATCACTTGAAGAGTCAAGACGAGCTGG 1252
    || || || || || || || || || || || || || || || || || || ||
DB 12018 GCACCTTTGGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 12077
QY 1253 CCAAATGCTGAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1312
    || || || || || || || || || || || || || || || || || || ||
DB 12078 CCAATATGTGAATCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 12137
QY 1313 GCGCTGTATCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1372
    || || || || || || || || || || || || || || || || || || ||
DB 12138 GTCCCTGTATCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 12197
QY 1373 CAGAGGTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1432
    || || || || || || || || || || || || || || || || || || ||
DB 12198 CGGAGGTGCACTGAGCGGAGATCAACACCTATCTCTCTCTCTCTCTCTCTCTCT 12257
QY 1433 CTCTGTCTCAAAAAAAAAAAAAAAAAAAAA 1459
    || || || || || || || || || || || || || || || || || || ||
DB 12258 CTCAGTCTCAAAAAAAAAAAAAAAAAAAAA 12284

```

RESULT 12
 AC005674 220000 bp DNA HTG 16-JUN-1999
 LOCUS AC005674
 DEFINITION Homo sapiens chromosome 4, WORKING DRAFT SEQUENCE, 9 unordered
 pieces.
 AC005674
 NID 94803911
 VERSION AC005674.7 GI:4803911
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 220000)
 Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
 Direct Submission
 Unpublished
 2 (bases 1 to 220000)
 Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
 Direct Submission
 Submitted (12-SEP-1998) Department of Genetics, Stanford Human
 Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
 On May 12, 1999 this sequence version replaced gi:4757680.

COMMENT
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. The true order of the pieces


```

/note="match: proteins P87304 P39828 CE02501 CE00158
P39514; supported by GENSCAN and EGENES"
/codon_start=1
/evidence=not_experimental
/product="dj30M3.1 (novel protein similar to (predicted)
plant, worm, yeast and archaea bacterial proteins)"
/protein_id="CA21139.1"
/db_xref="PID:e1373175"
/db_xref="PID:g4210351"
/db_xref="GI:4210351"
/translation="ITLVSAAPGVKVICEMKVEEHNIAICTLGGTATLVNISTMA
LITCRGAPGVSDMTNTYMSAPKLEGDIVYTHAVLKQKRTLAFTSDTLNKKATGKLI
AQRHRTNLEN"
13295..13604
/note="AlusX repeat: matches 1..310 of consensus"
repeat_region
14521..14821
/note="Aluy8 repeat: matches 1..299 of consensus"
repeat_region
15123..15387
/note="AlusX repeat: matches 30..298 of consensus"
repeat_region
15514..15624
/note="AluJo repeat: matches 2..136 of consensus"
repeat_region
16957..17250
/note="AluSg repeat: matches 1..294 of consensus"
repeat_region
17326..17589
/note="MIR repeat: matches 194..258 of consensus"
repeat_region
17534..17637
/note="L2 repeat: matches 2588..2692 of consensus"
repeat_region
17943..18245
/note="Aluya5 repeat: matches 1..311 of consensus"
repeat_region
18551..18684
/note="L2 repeat: matches 2357..2495 of consensus"
repeat_region
18704..18998
/note="AlusX repeat: matches 6..300 of consensus"
repeat_region
19037..19404
/note="match: GSS B40022"
repeat_region
19504..19639
/note="MIR repeat: matches 88..238 of consensus"
repeat_region
20030..20303
/note="AlusX repeat: matches 24..297 of consensus"
repeat_region
misc_feature
complement(20331..20463)
/note="match: GSS B74972"
repeat_region
21207..21337
/note="MIR repeat: matches 90..244 of consensus"
repeat_region
21347..21600

Query Match      16.1%  Score 234.6;  DB 9;  Length 102200;
Best Local Similarity 77.7%;  Pred. NO. 6.7e-45;
Matches 296;  Conservative 0;  Mismatches 84;  Indels 1;  Gaps 1;

QY 1080 TGTATCCAGAGGAAACAGTGACACAGATTGATAGGTGGAACACGAAAGAAACAGGA 1139
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 83884 TCTAGCGTGGGTGACAGAGTGAGACTGTGTATAAAAAATATAATATATAGCGTGAC 83825

QY 1140 AAAATCAAGACTCTACAAAGAGCGTGGCAGGCGTGCATGCGTGAATCCAGCACTT 1199
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 83824 AAAAATCTACATTTAAAAAATGCGGGGTGCAGCGGCTCACACCTGTAATCCAGCACTT 83765

QY 1200 TGGGAGGC-GAGGCAGCGAGATCACTTAGAGTAAGAGATTCAAGACCGCCTGGCCAAA 1258
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 83764 TGGGAGCGTGGAGTGGGAGATCACTTAAGGTTCAGGAATTTGAGGCCAGTGTGGCAATA 83705

QY 1259 TGTGTAATCTGTGCTGACTACAAAAATACAAAGTACTGGATATGATGGCAGGCGCT 1318
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 83704 TGGTGAACCCCTTCTCTATTAAAAATATAAAAAATATAGCCAGGRTAGTGCATGCGCT 83645

QY 1319 GTAATCCAGCTACTTGGAGGCTGAGGCAGAGAAATTGCTTGAATATAGGAGCGAGAG 1378
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 83644 GTAATCCAGCTACTTGGAGGCTGAGGCAGAGAAATTGCTTGAACCCGGAGGCGAGAG 83585

QY 1379 TTGAAGTAGATTGAGATCACACCACATATACCTCCAGCTGGGCAACAGAGTAAAGCTCTGT 1438
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 83584 TTGCGATGAGCTGAGATTGACACCACTGTATCTTCAGTCTGGGTGAGACAGGCAAGATTCTGT 83525

```


repeat_region complement(14592..14896)
/rpt_family="AluSg"
repeat_region 15215..15420
/rpt_family="L1PA2"
repeat_region 15439..15570
/rpt_family="AluSc"
repeat_region 15576..15984
/rpt_family="L1PA2"
repeat_region 15985..16154
/rpt_family="AluSg/x"
repeat_region 16188..16423
/rpt_family="AluSg/x"
repeat_region complement(16845..17079)
/rpt_family="MER47A"
repeat_region 17137..17231
/rpt_family="MER4_internal"
repeat_region 17271..17756
/rpt_family="MER9"
repeat_region complement(17856..17990)
/rpt_family="MER85"
repeat_region 17994..18235
/rpt_family="MER4_internal"
repeat_region 18062..18793
/rpt_family="MER31_internal"
repeat_region 18244..18981
/rpt_family="MER4_internal"
repeat_region 19480..19549
/rpt_family="L2"
repeat_region complement(19734..19786)
/rpt_family="Alu_rich"
repeat_region 19857..20203
/rpt_family="AluJo"
repeat_region 20224..20473
/rpt_family="MER31_internal"
repeat_region 20527..20885
/rpt_family="MER65_internal"
repeat_region 20894..21195
/rpt_family="AluSx"
repeat_region 21235..21528
/rpt_family="MER65_internal"
repeat_region 21535..22886
/rpt_family="L1PA13"
repeat_region 23008..23307
/rpt_family="AluSc"
repeat_region 23338..23509
/rpt_family="MER65_internal"
repeat_region 23532..23975
/rpt_family="MLT2A"
repeat_region 23977..24473
/rpt_family="MER65_internal"
repeat_region 24474..24888
/rpt_family="MER65A"
repeat_region complement(25342..25682)
/rpt_family="MIR"
repeat_region 25986..26154
/rpt_family="L1R22"
repeat_region 26180..26394
/rpt_family="L1R22"
repeat_region complement(26739..27076)
/rpt_family="MER7A"

Query Match 16.1% Score 234.6; DB 11; Length 80507;
Best Local Similarity 85.8%; Pred. No. 6.6e-45;
Matches 272; Conservative 0; Mismatches 44; Indels 1; Gaps 1;

QY 1143 ATCAAGACTCTACAAAGAGCGTGGCGTCTGCTGCTTAATCCAGACATTGG 1202
DB 4001 ATAAATAATTATCAATCTGGGCGGCGAGTGCCTCATGCTCAATCCAGACATTGG 4060
QY 1203 GAGGCG-AGGCGAGCAGATCACTTGAAGTAAAGATTCAAGACCAGCTGGCCAAATGG 1261
DB 4061 GAGGCGAGGTGGCGAGTCACTGAGTTCAGAGAGTTGAGACCAGCTGGCCAAATGG 4120

QY 1262 TGAATCTGTCTGTACTAAAAATCAAAAGTAGCTGATATGCTGAGCGCCCTGA 1321
DB 4121 CAAAACCCGTCTCTATTAAAAATTAAGTGTGCGATGTCAGACGCGCTGA 4180
QY 1322 ATCCAGCTACTTGGAGGCTGAGGAGAGAAATTGCTGAATATGGAGGAGATTG 1381
DB 4181 ATCCAGCTACTGCGGAGGCTGAGGAGAGATTCCTGAACCTGGGAGGAGATTG 4240
QY 1382 AAGTAGTTGAGATCAACCACTATCTCAGCTGGGCAACAGATGAAGCTGTCTC 1441
DB 4241 TAGTAGCTGAGATTCACCACTGCTCAGCCCTGGCAACAGATGAAGCTGTCTC 4300
QY 1442 AAAAAAAAAAAAAAAAAA 1458
DB 4301 AAAAAAAAAAAAAAAAAA 4317

RESULT 15
AC004971
LOCUS Homo sapiens clone DJ1125K23, WORKING DRAFT SEQUENCE, 21 unordered
DEFINITION pieces.
ACCESSION AC004971
NID 93213026
VERSION AC004971.1 GI:3213026
KEYWORDS HTG, HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 188477)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 188477)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1820: contig of 1820 bp in length
* 1821 1839: gap of unknown length
* 1840 3311: contig of 1472 bp in length
* 3312 3330: gap of unknown length
* 3331 5118: contig of 1788 bp in length
* 5119 5137: gap of unknown length
* 5138 6628: contig of 1491 bp in length
* 6629 6647: gap of unknown length
* 6648 8860: contig of 2213 bp in length
* 8861 8879: gap of unknown length
* 8880 10635: contig of 1756 bp in length
* 10636 10655: gap of unknown length
* 10656 13630: contig of 2976 bp in length
* 13631 13649: gap of unknown length
* 13650 17252: contig of 3603 bp in length
* 17253 20154: contig of 2883 bp in length
* 20155 20173: gap of unknown length
* 20174 24721: contig of 4548 bp in length
* 24722 24740: gap of unknown length
* 24741 30872: contig of 6132 bp in length
* 30873 30891: gap of unknown length
* 30892 36985: contig of 6094 bp in length
* 36986 37004: gap of unknown length

```
*      37005      40411: contig of 3407 bp in length
*      40412      40430: gap of unknown length
*      40431      50114: contig of 9684 bp in length
*      50115      50133: gap of unknown length
*      50134      67886: contig of 17753 bp in length
*      67887      67905: gap of unknown length
*      67906      84164: contig of 16259 bp in length
*      84165      84183: gap of unknown length
*      84184      105583: contig of 21400 bp in length
*      105584      105602: gap of unknown length
*      105603      128035: contig of 22434 bp in length
*      128037      128055: gap of unknown length
*      128056      185122: contig of 57067 bp in length
*      185123      185140: gap of unknown length
*      185141      186840: contig of 1700 bp in length
*      186841      186858: gap of unknown length
*      186859      188477: contig of 1619 bp in length.
```

FEATURES
Location/Qualifiers
1. 188477

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DJ1125K23"

BASE COUNT 43810 a 48919 c 48642 g 46728 t 378 others
ORIGIN

Query Match 16.1%: Score 234.6; DB 34; Length 188477;
Best Local Similarity 84.3%: Pred. No. 7.1e-45;
Matches 264; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

```
OY 1146 AAGACTCTACAAGAGCGCTGGCAGGTGCTCATGCTGTATCCAGCACTTGGGAG 1205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 86621 AGGATATGCAAAAGGCGTGGGACAGTGGGCTGACCTGTATCCAGCACTTGGAG 86680

OY 1206 GCGAGGCAAGCAGATCACTTGAAGTAAGAGTTCAAGACCAAGCTGGCCAAATGCTGA 1265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 86681 CCAAGGCGGCAATCACTTGGGAGTTCAAGACCAAGCTGGGCAACATGCTGAA 86740

OY 1266 ATCTGTCTACTAAAATCAAAAGTTAGTGATGATGGGCAAGCGCCTGTATCC 1325
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 86741 ACCCTGTCTACTAAAATCAAAATTAATCTGAGCGTGGGCGGCACTGTATCC 86800

OY 1326 CAGCTACTTGGAGGCTGAGCAGAGATTGCTGAATATGGAGGCAAGGTTGAAGT 1385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 86801 CAGCTACTTGAAGGCTGAGGAGAGATGCTGTAGCCTGGGAAGTGGAGTTGTGT 86860

OY 1386 GAGTTGAGATCACACCACTATCTCAGCTGGGCAACAGAGTAAGTCTGTCTCAAA 1445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 86861 GAGCGAGATGCGGCACTGCACTCAGCCTGGGCAAGAGTGAAGTCTGTCTCAAA 86920

OY 1446 AAAAAAAAAAAAAA 1458
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 86921 AAAAAAAAAAAAAA 86933
```

Search completed: September 25, 1999, 23:36:02
Job time: 7305 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 09:56:30 ; Search time 273 seconds
(without alignments)
1337.107 Million cell updates/sec

Title: US-09-030-606-174
Perfect score: 1459
Sequence: 1 GGTACAGCCGACACTGTTC.....TCAAAAAAAAAAAAAAAAAAAA 1459

Scoring table: IDENTITY_NNC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1456	99.8	1459	1	V58646
2	1456	99.8	1459	1	V61251
3	474	32.5	1265	1	V58645
4	474	32.5	1265	1	V61250
5	329	22.5	1386	1	V1855
6	329	22.5	1248	1	V58644
7	329	22.5	1248	1	V61249
8	292.8	20.1	1167	1	V58647
9	292.8	20.1	1167	1	V61252
10	291.8	20.0	871	1	V37495
11	289.4	19.8	402	1	X41114
12	257.4	17.6	1119	1	V58648
13	257.4	17.6	1119	1	V61253
14	226.2	15.5	24026	1	T18325
15	226.2	15.5	24026	1	T17455
16	226.2	15.5	24026	1	T17512
17	226.2	15.5	24026	1	T17513
18	226.2	15.5	24026	1	T17514
19	226.2	15.5	24026	1	T17515
20	226.2	15.5	24026	1	T17516
21	226.2	15.5	24026	1	T17517
22	226.2	15.5	24026	1	T17518
23	226.2	15.5	24026	1	T17519
24	226.2	15.5	24026	1	T17520
25	226.2	15.5	24026	1	T17521
26	226.2	15.5	24026	1	T17522
27	226.2	15.5	24026	1	T17523
28	226.2	15.5	24026	1	T17524
29	226.2	15.5	24031	1	T17525
30	226.2	15.5	24026	1	T17526
31	226.2	15.5	24026	1	T17527
32	226.2	15.5	24026	1	T17528
33	226.2	15.5	24026	1	T17529
34	226.2	15.5	24026	1	T17530
35	226.2	15.5	24026	1	T32612
36	223.2	15.3	50000	1	X23517
37	219.2	15.0	7676	1	V62906
38	218.6	14.9	7849	1	O94109
39	216.8	14.9	452	1	T42809
40	215.4	14.8	1618	1	O10207
41	215.4	14.8	1618	1	O46958
42	215.4	14.8	3234	1	O92781
43	213.4	14.6	20303	1	T71699

ALIGNMENTS

RESULT 1
ID V58646 standard; CDNA: 1459 BP.
AC V58646;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone D6.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy: ss.
OS Homo sapiens.
PN W09837418-A2.
PD 27-AUG-1998.
PE 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR MPI: 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful for detecting and treating prostate cancers
PS Claim 1; Page 114; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can be used in the method of the invention. The method is for detecting prostate cancer comprising contacting a biological sample with an agent able to bind an immunogenic portion of a prostate protein (such as encoded by this sequence). An antibody which binds to an immunogenic portion of the prostate protein, and the method can be used to detect, monitor progression of, or treat prostate cancers. The antibody may also be conjugated to a therapeutic agent for use in therapy of prostate cancers.
SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Query Match 99.8%; Score 1456; DB 1; Length 1459;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGTACAGCCGACACTGTTCAGAGATGAGTGCAGAGCTCTACACATCGGGCTGGCC 60
|||
DB 1 GGTACAGCCGACACTGTTCAGAGATGAGTGCAGAGCTCTACACATCGGGCTGGCC 60
QY 61 TGCACAGTCTTGAGGCGGACCAAGAGCCGAGAGTGGTGGAGGCGAGCCCTCCG 120
|||
DB 61 TGCACAGTCTTGAGGCGGACCAAGAGCCGAGAGTGGTGGAGGCGAGCCCTCCG 120
QY 121 TACGGACCCAGAGTACACAGAGACCTTGTCTGCTACAGCACTCATCATCAAGTTGG 180
|||
DB 121 TACGGACCCAGAGTACACAGAGACCTTGTCTGCTACAGCACTCATCATCAAGTTGG 180
QY 121 TACGGACCCAGAGTACACAGAGACCTTGTCTGCTACAGCACTCATCATCAAGTTGG 180
|||
DB 121 TACGGACCCAGAGTACACAGAGACCTTGTCTGCTACAGCACTCATCATCAAGTTGG 180
QY 181 ACGAATCCGTGTCGAGTGTACACCATCCGAGCATACACATGCTTGGCACTGCCCTA 240
|||
DB 181 ACGAATCCGTGTCGAGTGTACACCATCCGAGCATACACATGCTTGGCACTGCCCTA 240
QY 241 CCGGGGGAACCTTCTGCTGCTTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 300
|||
DB 241 CCGGGGGAACCTTCTGCTGCTTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 300
QY 301 GTGTGTGTGTCCTCTTCAAGAGAGTCTCTGCCCAGTGCAGGAGGCTGACCAAGCT 360
|||
DB 301 GTGTGTGTGTCCTCTTCAAGAGAGTCTCTGCCCAGTGCAGGAGGCTGACCAAGCT 360
QY 361 CTGCGTCCAGGAGCAATGCTTACCTGCTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 420
|||
DB 361 CTGCGTCCAGGAGCAATGCTTACCTGCTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 420
QY 421 NGAGTCTGAGTAAAGCTATGAGACCCGCTGATACACCCCAACATTTCTGGCGCGGCG 480
|||
DB 421 NGAGTCTGAGTAAAGCTATGAGACCCGCTGATACACCCCAACATTTCTGGCGCGGCG 480

```
|||||
Db 421 NCAGGCTCATTAAGCTCTATGACCCGCTGTACACCCCAACATCTTCTGGCCGGCG 480
QY 481 AGGGCAAGACAGAGAGACTCTCGCAAGCTGAGAGAGGGGAAAGGGAGGCGGAGCT 540
Db 481 AGGGCAAGACAGAGAGACTCTCGCAAGCTGAGAGAGGGGAAAGGGAGGCGGAGCT 540
QY 541 CAGGGAAGGTTGAGAGAGGGGAGACAGACACAGGGGCGCATGGGAGATGACAG 600
Db 541 CAGGGAAGGTTGAGAGAGGGGAGACAGACACAGGGGCGCATGGGAGATGACAG 600
QY 541 CAGGGAAGGTTGAGAGAGGGGAGACAGACACAGGGGCGCATGGGAGATGACAG 600
Db 541 CAGGGAAGGTTGAGAGAGGGGAGACAGACACAGGGGCGCATGGGAGATGACAG 600
QY 601 ATGAGAGACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 601 ATGAGAGACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 661 ATAAACACAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Db 661 ATAAACACAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 721 AGAAACACACAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Db 721 AGAAACACACAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 781 GACCTCCACCAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Db 781 GACCTCCACCAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
QY 841 ATAGCTTCTGTTGACAGGGAGGCTTACCAATACATTAATAGTCAATTTATGAC 900
Db 841 ATAGCTTCTGTTGACAGGGAGGCTTACCAATACATTAATAGTCAATTTATGAC 900
QY 901 TTATGCTTATGATATATACCTTGTGGAATTTTGTGATATTTTGAAGTACAGG 960
Db 901 TTATGCTTATGATATATACCTTGTGGAATTTTGTGATATTTTGAAGTACAGG 960
QY 961 GCTGTGAAATTTTAAATGTTGCAACTCTCTAAATTTTTCGATGTTTATGTA 1020
Db 961 GCTGTGAAATTTTAAATGTTGCAACTCTCTAAATTTTTCGATGTTTATGTA 1020
QY 1021 AAAAATCCAGTTAAGTGGACTGTGCAATTCAAACAGGGTGTTCAGGGTCAACTGT 1080
Db 1021 AAAAATCCAGTTAAGTGGACTGTGCAATTCAAACAGGGTGTTCAGGGTCAACTGT 1080
QY 1081 GTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Db 1081 GTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
QY 1141 AAATCAAGACTCTACAAAGAGGCTGGGCAAGGCTGCTATGCTGTATATCCAGCACTTT 1200
Db 1141 AAATCAAGACTCTACAAAGAGGCTGGGCAAGGCTGCTATGCTGTATATCCAGCACTTT 1200
QY 1201 GGGAGGCGAGGCGAGAGATCACTTGAAGTAAAGAGTTCAAGACAGAGCTGGGCAAAATG 1260
Db 1201 GGGAGGCGAGGCGAGAGATCACTTGAAGTAAAGAGTTCAAGACAGAGCTGGGCAAAATG 1260
QY 1261 GTGAATTCCTGTCTACTATAAAATACAAAGTTAGCTGATATGTTGGAGGCGCTGT 1320
Db 1261 GTGAATTCCTGTCTACTATAAAATACAAAGTTAGCTGATATGTTGGAGGCGCTGT 1320
QY 1321 AATCCAGCTACTTGGAGGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Db 1321 AATCCAGCTACTTGGAGGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1381 GAAGTAGTTGAGATACACACTATATCTCAAGTGGGCAACAGAGTAAGACTGTCT 1440
Db 1381 GAAGTAGTTGAGATACACACTATATCTCAAGTGGGCAACAGAGTAAGACTGTCT 1440
QY 1441 CAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1459
Db 1441 CAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1459
```

RESULT - 2*

```
V61251
ID V61251 standard; cDNA; 1459 BP.
AC V61251.
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE6.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN W09837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; U020956.
PR 25-FEB-1997; U0806099.
PR 01-AUG-1997; U0904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J.
DR WPI, 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3: Page 106; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Query Match 99.8%; Score 1456; DB 1; Length 1459;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTGACGCCGACACTGTTTCCAGAAAGTGAAGTGCACAGCTCTTACACCATCGGGCTGGGCC 60
Db 1 GGTGACGCCGACACTGTTTCCAGAAAGTGAAGTGCACAGCTCTTACACCATCGGGCTGGGCC 60
QY 61 TGCACAGTCTTGAAGCCGACCAAGAGCCAGGAGCCAGATGTTGAGAGGCGACCTCTCCG 120
Db 61 TGCACAGTCTTGAAGCCGACCAAGAGCCAGGAGCCAGATGTTGAGAGGCGACCTCTCCG 120
QY 121 TACGGCACCAGAGTACACAGAGACCCTTGCTGCTAGCAGCTCATCAAGTTGG 180
Db 121 TACGGCACCAGAGTACACAGAGACCCTTGCTGCTAGCAGCTCATCAAGTTGG 180
QY 181 AGGAATCCGTGTCAGAGTGTGACACCATCCGAGATAGCATTTGCTTGGAGTCCCTA 240
Db 181 AGGAATCCGTGTCAGAGTGTGACACCATCCGAGATAGCATTTGCTTGGAGTCCCTA 240
QY 241 CCGGGGGAACCTTGGCTCGTTTGGCTGGGAGTCTGCTGGAGAGAGGTAGCTACAGG 300
Db 241 CCGGGGGAACCTTGGCTCGTTTGGCTGGGAGTCTGCTGGAGAGAGGTAGCTACAGG 300
QY 301 GTGTGTCTGCTCCTTTCAGAGAGGTCCTTGCCAGTGGCGGGGCTGACCCAGAGCT 360
Db 301 GTGTGTCTGCTCCTTTCAGAGAGGTCCTTGCCAGTGGCGGGGCTGACCCAGAGCT 360
QY 361 CTGCGTCCAGGAGAGATCCCTACAGTGTGAGTGGAGTGGAGTGGTGGTGTCTGA 420
Db 361 CTGCGTCCAGGAGAGATCCCTACAGTGTGAGTGGAGTGGAGTGGTGGTGTCTGA 420
QY 421 NCAGGCTCATTAAGCTCTATGACCCGCTGTACACCCCAACATCTTCTGGCCGGCG 480
Db 421 NCAGGCTCATTAAGCTCTATGACCCGCTGTACACCCCAACATCTTCTGGCCGGCG 480
QY 481 AGGGCAAGACAGAGAGACTCTCGCAAGCTGAGAGAGGGGAAAGGGAGGCGGAGCT 540
Db 481 AGGGCAAGACAGAGAGACTCTCGCAAGCTGAGAGAGGGGAAAGGGAGGCGGAGCT 540
QY 541 CAGGGAAGGTTGAGAGAGGGGAGACAGACACAGGGGCGCATGGGAGATGACAG 600
Db 541 CAGGGAAGGTTGAGAGAGGGGAGACAGACACAGGGGCGCATGGGAGATGACAG 600
QY 601 ATGAGAGACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
```



```

|||||
Db 601 ATGAGAGACACACAGGAGACACTGACACTAGAGAGAGAACTGAGAGAAAGAGAA 660
OY 661 ATAAACAGAGAAATAAAGAGAAAGAAAGAGAGAAAGAGAAAGAGAAAGAGAAAGAGAA 720
Db 661 ATAAACAGAGAAATAAAGAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAA 720
OY 721 AGAAACACACACACATAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Db 721 AGAAACACACACACATAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
OY 781 GACCTCCACACATAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Db 781 GACCTCCACACATAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
OY 841 ATAGCCTACTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
Db 841 ATAGCCTACTGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
OY 901 TTATGCTATCATGATATATACCTTTGTTGAGATTTTGTATTTTGTATTTTGTATTTTGTATTT 960
Db 901 TTATGCTATCATGATATATACCTTTGTTGAGATTTTGTATTTTGTATTTTGTATTTTGTATTT 960
OY 961 GTCCTGATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 1020
Db 961 GTCCTGATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 1020
OY 1021 AAAATGCAAGTAAATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
Db 1021 AAAATGCAAGTAAATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
OY 1081 GTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Db 1081 GTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
OY 1141 AAATCAAGACTCTACAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200
Db 1141 AAATCAAGACTCTACAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200
OY 1201 GGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Db 1201 GGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
OY 1261 GTGAAATCTGCTGTCTGTAATAAATACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Db 1261 GTGAAATCTGCTGTCTGTAATAAATACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
OY 1321 AATCCAGACTGTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
Db 1321 AATCCAGACTGTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
OY 1381 GAAGTGAGTGAATTCACACACATATCTCAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 1381 GAAGTGAGTGAATTCACACACATATCTCAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAG 1440
OY 1441 CAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1459
Db 1441 CAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1459

```

RESULT 3

```

V58645
ID V58645 standard: cDNA: 1265 BP.
AC V58645;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE2.
KM Prostate tumour specific gene; human; prostate cancer; detection;
OS Homo sapiens.
PN W09837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998: U03690.
PR 09-FEB-1998: US-904809.

```

```

PR 25-FEB-1997: US-806596.
PR 01-AUG-1997: US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J.
DR WPI: 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1: Page 113-114; 14pp: English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1265 BP; 256 A; 432 C; 321 G; 245 T;

```

```

Query Match 32.5%; Score 474; DB 1; Length 1265;
Best Local Similarity 98.4%; Pred. No. 4,4e-95;
Matches 477; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

OY 25 AGTGAAGTGCAGAGCTCTACACCATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 84
Db 104 ACTGTTCCAGAACTCTACACCATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 163
OY 85 AGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 144
Db 164 AGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 223
OY 145 CTTGCTGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 204
Db 224 CTTGCTGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 283
OY 205 CCATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 264
Db 284 CCATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 343
OY 265 CTGAGTGGGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 324
Db 344 CTGAGTGGGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 403
OY 325 GGTCTCTGCGCCAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 384
Db 404 GGTCTCTGCGCCAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 463
OY 385 CGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 444
Db 464 CGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 523
OY 445 CCGGCTGTACACCCAGCATGTTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 504
Db 524 CCGGCTGTACACCCAGCATGTTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 583
OY 505 CAACG 509
Db 584 CAACG 588

```

RESULT 4

```

V61250
ID V61250 standard: cDNA: 1265 BP.
AC V61250;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE2.
KM Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN W09837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998: U03492.
PR 09-FEB-1998: US-020956.

```

PR 25-FEB-1997; US-806099.
 PR 01-AUG-1997; US-904804.
 PA (CORI-) CORIXA CORP.
 PI DILLON DC, Xu J.
 DI WPI: 98-609886/51.
 DR Polypeptides comprising immunogenic portions of prostate proteins -
 PT used in a vaccine for the treatment of prostate cancer
 PS Claim 3; Page 105-106; 130pp; English.
 CC The present sequence is a new DNA which encodes an immunogenic portion
 CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
 CC can be used as a vaccine for the treatment of prostate cancer. The DNA
 CC was identified by analysis of a subtracted cDNA library obtained by
 CC subtracting a prostate tumour cDNA expression library with a normal
 CC tissue cDNA library.
 SQ Sequence 1265 BP; 256 A; 432 C; 321 G; 245 T;

Query Match 32.5%; Score 474; DB 1; Length 1265;
 Best Local Similarity 98.4%; Pred. No. 4.4e-95;
 Matches 477; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 25 AATGAGTGAAGAGCTCTACACCATGGGCTGGCTGCACAGTCTTGAGGCCGACCAAG 84
 DB 104 AATGTTCCAGACTCTACACCATGGGCTGGCTGCACAGTCTTGAGGCCGACCAAG 163
 QY 85 AGCCAGGAGAGAGATGTGAGAGCCAGCTCTCCGACGACACCAAGATCAACAGAC 144
 DB 164 AGCCAGGAGAGAGATGTGAGAGCCAGCTCTCCGACGACACCAAGATCAACAGAC 223
 QY 145 CCTTGCTGCTAAGACACTCATGCTCATCAAGTTGAGACGAATCCGTGCGAGTGTGACA 204
 DB 224 CCTTGCTGCTAAGACACTCATGCTCATCAAGTTGAGACGAATCCGTGCGAGTGTGACA 283
 QY 205 CCATCCGAGAGATGAGATCTGCTGCACTGCTGCACTGCTGCACTGCTGCACTGCTGCACT 264
 DB 284 CCATCCGAGAGATGAGATCTGCTGCACTGCTGCACTGCTGCACTGCTGCACTGCTGCACT 343
 QY 265 CTGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 324
 DB 344 CTGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 403
 QY 325 GGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 384
 DB 404 GGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 463
 QY 385 CGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 444
 DB 464 CGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 523
 QY 445 CCGGCTGTACCAACCCATGTTCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 504
 DB 524 CCGGCTGTACCAACCCATGTTCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 583
 QY 505 CAACG 509
 DB 584 CAACG 588

RESULT 5
 VIL1855
 ID VIL1855 standard; cDNA: 1386 BP.
 AC VIL1855;
 DT 11-SEP-1998 (first entry)
 DE Homo sapiens Tub Interactor (hTI-1) gene.
 KW serine protease; tub interactor; treatment; obesity; cachexia;
 KW anorexia nervosa; diabetes; cell cycle progression; apoptosis;
 KW neurodegenerative disease; Alzheimer's disease; drug screening;
 KW Parkinson's disease; Huntington's chorea; detection; diagnosis;
 KW amyotrophic lateral sclerosis; spinocerebellar degeneration; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 2..701
 FT /tag= a

/product= hTI-1 protein
 /note= "putative serine protease"

FT M09812302-A1.
 FN 26-MAR-1998.
 PD 05-SEP-1997; U15627.
 PR 21-JUL-1997; US-897340.
 PR 17-SEP-1996; US-715032.
 PA (MILL-) MILLENNIUM PHARM INC.
 PI Errata PR, Gimenio CJ;
 DI WPI: 98-217246/19.
 DR P-PSDB; W59129.
 PT Tub interactor genes - used to develop products for the treatment
 PT of obesity, cachexia, anorexia nervosa or related disorders e.g.
 PT diabetes
 PS Claim 10; Fig 1; 120pp; English.
 CC The sequence is that of the tub interactor gene hTI-1 which
 CC codes for a putative serine protease. 71 genes function
 CC in biochemical pathways involved in weight control and
 CC related disorders. The products can be used for treating
 CC weight disorders, e.g. obesity, cachexia or anorexia nervosa,
 CC or a related disorder such as diabetes. The products can
 CC also be used to modulate cell cycle progression and apoptosis.
 CC They can be used for treating neurodegenerative diseases
 CC which are characterised by apoptosis, including Alzheimer's
 CC disease, Parkinson's disease, Huntington's chorea, amyotrophic
 CC lateral sclerosis or spinocerebellar degenerations. The
 CC products can also be used for detection, diagnosis and
 CC drug screening.
 SQ Sequence 1386 BP; 318 A; 490 C; 321 G; 249 T;

Query Match 22.5%; Score 329; DB 1; Length 1386;
 Best Local Similarity 83.1%; Pred. No. 2.3e-63;
 Matches 422; Conservative 0; Mismatches 3; Indels 83; Gaps 1;

QY 2 GTACCCGACACACTGTTTCCAGAAAGTGCAGAGCTCTACACCATGGGCTGGGCT 61
 DB 100 GTACCCGACACACTGTTTCCAGAAAGTGCAGAGCTCTACACCATGGGCTGGGCT 159
 QY 62 GCACAGTCTTGAAGGCGGACCAAGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 121
 DB 160 GCACAGTCTTGAAGGCGGACCAAGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 219
 QY 122 AGGAGACCAAGATGACACAGACCTTCTGCTAAGACCTCATGCTCATCAAGTTGGA 181
 DB 220 AGGAGACCAAGATGACACAGACCTTCTGCTAAGACCTCATGCTCATCAAGTTGGA 279
 QY 182 CGAATCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 241
 DB 280 CGAATCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 339
 QY 242 CCGGAG 301
 DB 340 CCGGAG 386
 QY 302 TGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 361
 DB 386 -----
 QY 362 TGCCTCCAGAGAGATGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 421
 DB 386 -----GGCAGATGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 436
 QY 422 GAGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 481
 DB 437 GAGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 496
 QY 482 GGGCAAGACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 509
 DB 497 GGGCAAGACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 524

RESULT 6

V58644
ID V58644 standard; cDNA; 1248 BP.
AC V58644;
DE 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DEL
KM Prostate tumour specific gene; human; prostate cancer; detection;
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 217..696
FT /tag- a
PN MO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillion DC, Xu J;
DR WPI: 98-480805/41.
DR P-PSDB: W69387.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 112; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 22.5%; Score 329; DB 1; Length 1248;
Best Local Similarity 83.1%; Pred. No. 2.2e-63;
Matches 422; Conservative 0; Mismatches 3; Indels 83; Gaps 1;

QY 2 GTACAGCCGACACGCTTTCCAGAAAGTGAAGTCAAGTCTCTACACCATCGGGCTGGGCT 61
DB 120 GTACAGCCGACACGCTTTCCAGAAAGTGAAGTCAAGTCTCTACACCATCGGGCTGGGCT 179
QY 62 GCAAGTCTTGAAGCCGACCAAGAGCCAGGAGCCAGTGTGAGGCGACCTCTCGCT 121
DB 180 GCAAGTCTTGAAGCCGACCAAGAGCCAGGAGCCAGTGTGAGGCGACCTCTCGCT 239
QY 122 ACGGACCCAGAGTACACAGACACCTTCTGCTAAGCACTGCTCATCAAGTTGGA 181
DB 240 ACGGACCCAGAGTACACAGACACCTTCTGCTAAGCACTGCTCATCAAGTTGGA 299
QY 182 GCAATCGGTGTCCAGTGTGACACCATCCGAGCATCGATGCTTGTGCGAGTGCCTAC 241
DB 300 GCAATCGGTGTCCAGTGTGACACCATCCGAGCATCGATGCTTGTGCGAGTGCCTAC 359
QY 242 CGCGGGGAACCTTGTGCTTGTGCTGGGTGTGCTGGCGAAGCGGTGAGCTACAGGG 301
DB 360 CGCGGGGAACCTTGTGCTTGTGCTGGGTGTGCTGGCGAAGCGGTGAGCTACAGGG 406
QY 302 TGTGTGTCTGCCCTTCCAGAGAGTCTGTGCCAGTGGGGGGGCTGACCCAGAGCTC 361
DB 406 ----- 406
QY 362 TGCCTCCAGGACAGATGCTACCGTGTGCTGACAGTGTGTAAGTGTGCTGTGTA 421
DB 406 -----GGCAGAAATGCTTACCGTGTGCTGCTGTAAGTGTGTAAGTGTGTA 456
QY 422 GAGGTGTGCAATTAAGCTTATGAGCCGCTGTACCAACCCCAACATGTTTCTGCGCGGCGGA 481
DB 457 GAGGTGTGCAATTAAGCTTATGAGCCGCTGTACCAACCCCAACATGTTTCTGCGCGGCGGA 516
QY 482 GGGCAAGACCAAGAGACTCTGCAACG 509

DB 517 GGGCAAGACCAAGAGACTCTGCAACG 544
RESULT 7
ID V61249
AC V61249;
DE 06-JUN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DEL
KM Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN MO9837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillion DC, Xu J;
DR WPI: 98-609886/51.
DR P-PSDB: W71871.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 104; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 22.5%; Score 329; DB 1; Length 1248;
Best Local Similarity 83.1%; Pred. No. 2.2e-63;
Matches 422; Conservative 0; Mismatches 3; Indels 83; Gaps 1;

QY 2 GTACAGCCGACACGCTTTCCAGAAAGTGAAGTCAAGTCTCTACACCATCGGGCTGGGCT 61
DB 120 GTACAGCCGACACGCTTTCCAGAAAGTGAAGTCAAGTCTCTACACCATCGGGCTGGGCT 179
QY 62 GCAAGTCTTGAAGCCGACCAAGAGCCAGGAGCCAGTGTGAGGCGACCTCTCGCT 121
DB 180 GCAAGTCTTGAAGCCGACCAAGAGCCAGGAGCCAGTGTGAGGCGACCTCTCGCT 239
QY 122 ACGGACCCAGAGTACACAGACACCTTCTGCTAAGCACTGCTCATCAAGTTGGA 181
DB 240 ACGGACCCAGAGTACACAGACACCTTCTGCTAAGCACTGCTCATCAAGTTGGA 299
QY 182 GCAATCGGTGTCCAGTGTGACACCATCCGAGCATCGATGCTTGTGCGAGTGCCTAC 241
DB 300 GCAATCGGTGTCCAGTGTGACACCATCCGAGCATCGATGCTTGTGCGAGTGCCTAC 359
QY 242 CGCGGGGAACCTTGTGCTTGTGCTGGGTGTGCTGGCGAAGCGGTGAGCTACAGGG 301
DB 360 CGCGGGGAACCTTGTGCTTGTGCTGGGTGTGCTGGCGAAGCGGTGAGCTACAGGG 406
QY 302 TGTGTGTCTGCCCTTCCAGAGAGTCTGTGCCAGTGGGGGGGCTGACCCAGAGCTC 361
DB 406 ----- 406
QY 362 TGCCTCCAGGACAGATGCTACCGTGTGCTGACAGTGTGTAAGTGTGCTGTGTA 421
DB 406 -----GGCAGAAATGCTTACCGTGTGCTGCTGTAAGTGTGTAAGTGTGTA 456
QY 422 GAGGTGTGCAATTAAGCTTATGAGCCGCTGTACCAACCCCAACATGTTTCTGCGCGGCGGA 481
DB 457 GAGGTGTGCAATTAAGCTTATGAGCCGCTGTACCAACCCCAACATGTTTCTGCGCGGCGGA 516
QY 482 GGGCAAGACCAAGAGACTCTGCAACG 509
DB 517 GGGCAAGACCAAGAGACTCTGCAACG 544

ID	Accession	Standard	CDNA	BP
AC	V58647	standard	CDNA	1167 BP
DT	V58647			
DE	08-DEC-1998	(first entry)		
KW	Prostate tumour specific gene clone DE13.			
OS	Prostate tumour specific gene; human; prostate cancer; detection;			
FW	therapy; ss.			
FN	Homo sapiens.			
FT	Key	Location/Qualifiers		
FT	CDS	28..645		
FT		/*tag= a		
PN	W09837418-A2.			
PD	27-AUG-1998.			
PR	25-FEB-1998; U03690.			
PR	09-FEB-1998; US-904809.			
PR	25-FEB-1997; US-806596.			
PA	01-AUG-1997; US-904809.			
PI	(CORI-) CORIXA CORP.			
PI	Dillon DC, Xu J;			
PI	WPI: 98-480805/41.			
PR	P-PSDB: W69388.			
PT	Novel human prostate specific tumour protein and fragments - useful			
PT	for detecting and treating prostate cancers			
PS	Claim 1; Page 115; 141pp: English.			
CC	This sequence represents a human prostate tumour specific gene, and can			
CC	be used in the method of the invention. The method is for detecting			
CC	prostate cancer comprises contacting a biological sample with an agent			
CC	able to bind an immunogenic portion of a prostate protein (such as			
CC	encoded by this sequence). An antibody which binds to an immunogenic			
CC	portion of the prostate protein, and the method can be used to detect,			
CC	monitor progression of, or treat prostate cancers. The antibody may			
CC	also be conjugated to a therapeutic agent for use in therapy of prostate			
CC	cancers.			
SQ	Sequence	1167 BP;	242 A;	400 C; 287 G; 222 T;

Query Match	Best Local Similarity	Matches 390; Conservative	Score 292.8; DB 1; Length 1167;	20.1%; 80.4%; Pred. No. 1.8e-55;	0; Mismatches 12; Indels 83; Gaps 14;
Db 429	CCCCGTGACACACCCACACATGTTCTGGCGCGGCGGAGGACAGACAGAAAGACTCTG	488			
Qy 505	CAACG 509				
Db 489	CAACG 493				
RESULT 9					
ID V61252	standard; cDNA; 1167 BP.				
AC V61252;					
DE 06-JAN-1999	(first entry)				
DE cDNA sequence of prostate tumour clone P703 splice variant DE13.					
DE prostate; cancer; tumour; vaccine; immunogen; clone; ss.					
OS Homo sapiens.					
PN W09837093-A2.					
PD 27-AUG-1998.					
PF 25-FEB-1998; U03492.					
PR 09-FEB-1998; US-020956.					
PR 25-FEB-1997; US-806099.					
PR 01-AUG-1997; US-904804.					
PR (CORI-) CORIXA CORP.					
PI Dillon DC, Xu J.					
PI WPI; 98-609886/51.					
DR P-PSDB; W71782.					
PT Polypeptides comprising immunogenic portions of prostate proteins -					
PS used in a vaccine for the treatment of prostate cancer					
PS Claim 3; Page 107; 130pp; English.					
CC The present sequence is a new DNA which encodes an immunogenic portion					
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,					
CC can be used as a vaccine for the treatment of prostate cancer. The DNA					
CC was identified by analysis of a subtracted cDNA library obtained by					
CC subtracting a prostate tumour cDNA expression library with a normal					
CC tissue cDNA library.					
SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;					

ID	Accession	Location/Qualifiers
RESULT 10		
V37495		
V37495	standard; DNA: 871 BP.	
AC	V37495.	
DT	07-SEP-1998 (first entry)	
KD	Human prostate-specific kallikrein (HPSK) encoding DNA.	
DE	Prostate-specific kallikrein; HPSK; prostate carcinoma; human;	
KW	benign prostate hyperplasia; diagnosis; drug screening; PSK, ss.	
OS	Homo sapiens.	
FT	Key	
FT	Location/Qualifiers	
FT	31..777	
FT	/tag= a	
FT	/transl_except= (pos:367..369, aa:Xaa)	
FT	/transl_except= (pos:412..414, aa:Xaa)	
FT	/transl_except= (pos:424..426, aa:Xaa)	
FT	/product= "HPSK protein"	
FT	/note= "Xaa = unknown"	
PN	W09820117-A1.	
PD	14-MAY-1998.	
PE	31-OCT-1997; U20051.	
PR	05-NOV-1996; US-744026.	
PR	(INCY-) INCYTE PHARM INC.	
PI	Bandman O, Goli SK;	
PI	WPI; 98-286933/25.	
DR	P-PSDB: W60592.	
PT	New isolated prostate-specific kallikrein - used to develop products	
PT	for diagnosis and treatment of, e.g. prostate carcinoma or benign	
PT	hyperplasia	
PS	Claim 5; Fig 1A-C; 68pp; English.	
CC	This DNA encodes a human prostate-specific kallikrein (HPSK). A host cell	
CC	containing an expression vector comprising the HPSK nucleic acid sequence	
CC	can be used to produce the protein recombinantly. The HPSK products can	
CC	be used for the diagnosis of conditions or diseases associated with	
CC	expression of HPSK such as prostate carcinoma and benign prostate	
CC	hyperplasia. Agonists and antagonists which specifically bind to HPSK and	
CC	modulate its activity can be used for the preparation of treatment of	
CC	such conditions or diseases. The products can also be used for detection	
CC	and drug screening, especially for the detection of prostate-specific	
CC	kallikrein (PSK).	
SQ	Sequence 871 BP; 166 A; 260 C; 258 G; 184 T;	
Query Match	20.0%; Score 291.8; DB 1; Length 871;	
Best Local Similarity	80.2%; Pred. No. 2.7e-55;	
Matches 389; Conservative	0; Mismatches 13; Indels 83; Gaps 1.	
QY	25 AGTGTGATGACAGCTCTACACCATCGGGCTGGCCTGCACAGTCTTGGAGCCGACCAAG 84	
DB	224 ACTGTTCCAGAACCTCTACACCATCGGGCTGGCCTGCACAGTCTTGGAGCCGACCAAG 283	
QY	85 AGCCAGGAGCCAGATGTGTGGAGCCAGCTCTCCGTACGGACCCAGAGTCAACACAGC 144	
DB	284 AGCCAGGAGCCAGATGTGTGGAGCCAGCTCTCCGTACGGACCCAGAGTCAACACAGC 343	
QY	145 CCTTCTGGCTACGACCTCATGCTCATTAAGTTGGACCAATCCGTGTCCGAGTCTGACA 204	
DB	344 CCTTCTGGCTACGACCTCATGCTCATTAAGTTGGACCAATCCGTGTCCGAGTCTGACA 403	
QY	205 CCATCCGGAGCATTCAGATTCCTTGGACATGCGCTACCCGGGGAACTTTGCTCTGTTT 264	
DB	404 ACATCCGGAGCATTCAGATTCCTTGGACATGCGCTACCCGGGGAACTTTGCTCTGTTT 463	
QY	265 CTGGCTGGGGTCTGTGTGGGACGGGTGAGCTACGGGTGTGTCTGCCCTTCAAGA 324	
DB	464 CTGGCTGGGGTCTGTGTGGGACGGGTGAGCTACGGGTGTGTCTGCCCTTCAAGA 487	
QY	325 GGTCTCTGTGCTGAGTGGGGGGGCTGACCCAGAGCTCTGCTCCAGGAGAAATGCTTAC 384	

DB	487	-----	GGAGAAATGCGTAC	500
OY	385	CGTGTGCAAGTCGTCGAACGTCGTGGTGTCCTGANGAGTCTGCANTAACTCTATGA	444	
DB	501	CGTGTGTGAGTGTCCGTGAACGTCGTGGTGTCCTGANGAGTCTGCANTAACTCTATGA	560	
OY	445	CCCCGTGACCAACCCCAACATGTTCTGGCGGGGAGGAGGCAAGACCAAGAGCTCCG	504	
DB	561	CCCCGTGACCAACCCCAACATGTTCTGGCGGGGAGGAGGCAAGACCAAGAGCTCCG	620	
OY	505	CAACG 509		
DB	621	CAACG 625		
RESULT	11			
XA1114				
ID	XA1114 standard; cDNA; 402 BP.			
AC	XA1114:			
DT	17-JUN-1999 (first entry)			
DE	Human secreted protein 5', EST SEQ ID NO:58.			
KW	Human; secreted protein; EST; expressed sequence tag; diagnosis;			
KW	forensic; gene therapy; chromosome mapping; signal peptide;			
KW	upstream regulatory sequence; cytokine activity; cell proliferation;			
KW	differentiation; haematopoiesis regulation; tissue growth regulation;			
KW	reproductive hormone regulation; chemotactic; chemokine; haemostatic;			
OS	thrombolytic; anti-inflammatory; tumour inhibition; ds.			
PN	Homo sapiens.			
PD	MO9906548-A2.			
PI	11-FEB-1999.			
PR	31-JUL-1998; IB1222.			
PR	01-AUG-1997; US-905135.			
PA	(GEST) GENSET.			
PI	Duclet A, Dumas Mline Edwards J, Lacroix B;			
DR	WPI; 99-153778/13.			
DR	P-PSDB; Y12281			
PT	New nucleic acids encoding human secreted proteins - obtained from			
PT	cDNA libraries prepared from e.g. liver, ovary, brain, prostate,			
PT	kidney, lung, umbilical cord, placenta and colon tissue			
PS	Claim 1; Page 205; 824pp; English.			
CC	XA1094 to XA1347 represent 5' expressed sequence tags (ESTs) for human			
CC	secreted proteins, and encode the proteins given in Y12281 to Y12514,			
CC	respectively. The proteins given represent the signal peptide and an			
CC	N-terminal fragment of a secreted protein. The nucleic acid sequences			
CC	can be used for producing secreted human gene products. They can also			
CC	be used to develop products for diagnosis and therapy. The proteins			
CC	obtained may have cytokine activity, cell proliferation/differentiation			
CC	activity, haematopoiesis regulating activity, tissue growth regulating			
CC	activity, reproductive hormone regulating activity, chemotactic/			
CC	chemokine activity, haemostatic and thrombolytic activity, receptor/			
CC	ligand activity, anti-inflammatory activity, tumour inhibition activity			
CC	or other activities. The products can be used in forensic, gene therapy			
CC	and chromosome mapping procedures. The sequences can also be used for			
CC	obtaining corresponding promoter sequences. The nucleic acids encoding			
CC	the signal peptide can be used for directing extracellular secretion of			
CC	a polypeptide or the insertion of a polypeptide into a membrane, or			
CC	importing a polypeptide into a cell.			
SO	Sequence 402 BP; 80 A; 125 C; 121 G; 75 T;			
Query Match	19.8%; Score 289.4; DB 1; Length 402;			
Best Local Similarity	99.0%; Pmed. No. 7.6e-55;			
Matches 290; Conservative 1; Mismatches 2; Indels 0; Gaps 0;				
OY	2	GTACACGGCACAAGTTCAGAACTGATGAGAGGTCCTACCAATCGGCTGGGCT	61	
DB	108	GTACACGGCACAAGTTCAGAACTGATGAGAGGTCCTACCAATCGGCTGGGCT	167	
OY	62	GCACAGCTTTGAGCGGACCAAGAGCCAGAGGACGATGATGAGAGCCAGCTTCCT	121	
DB	168	GCACAGCTTTGAGCGGACCAAGAGCCAGAGGACGATGATGAGAGCCAGCTTCCT	227	
OY	122	ACGGCACCAGATGACAAGACCTTGTCTGCTAAGACCTCATGCTCATTAAGTTGA	181	

```
Db 228 ACGGACCCAGATACACAGACCTTGTCTGCTAAGACCTATGCTCAAGTTGA 287
|||||
QY 182 CGAATCGGTGCGAGCTGACACATCCGAGCATCAGATGCTTGGAGTCCCTAC 241
|||||
Db 288 CGATCCGAGTGTGAGTCTACACATCCGAGCATCAGATGCTTGGAGTCCCTAC 347
|||||
QY 242 CGCGGGGAATCTTGTCTGCTTGTGCTGGGCTGCTGCTGGCAACGGTGAGC 294
|||||
Db 348 CGCGGGGAATCTTGTCTGCTTGTGCTGGGCTGCTGCTGGCAACGGTGAGC 400
|||||

RESULT 12
V58648
ID V58648 standard; cDNA; 1119 BP.
AC V58648;
DE 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DEL14.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KM therapy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 34..528
FT /tag= a
FT
PN W09837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillion DC, Xu J;
DR WPI; 98-480805/41.
DR P-PSDB; W69389.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1: Page 116-117; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match 17.6%; Score 257.4; DB 1; Length 1119;
Best Local Similarity 97.8%; Pred. No. 9.6e-48;
Matches 261; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 25 AGTGAATGCAGACTCTTAACCATGCGGCTGGGCTGCACAGTCTTGAAGCCGACCAAG 84
|||||
Db 98 ACTGTTCCAGAACTCTTAACCATGCGGCTGGGCTGCACAGTCTTGAAGCCGACCAAG 157
|||||
QY 85 AGCCAGGAGACCGAGATGGTGGAGGCCCTCTCCGTAAGGAGCCCGAGATACAACAGAC 144
|||||
Db 158 AGCCAGGAGACCGAGATGGTGGAGGCCCTCTCCGTAAGGAGCCCGAGATACAACAGAC 217
|||||
QY 145 CCTTGTCTGCTAAGACCTCATGCTCATCAAGTTGAGAGATCCGTTCCGAGTGTGACA 204
|||||
Db 218 CCTTGTCTGCTAAGACCTCATGCTCATCAAGTTGAGAGATCCGTTCCGAGTGTGACA 277
|||||
QY 205 CCATCCGAGACATCAGCATTTGCTTGCAGTCCCTACCGCGGGGAATCTTGCCTGTTT 264
|||||
Db 278 CCATCCGAGACATCAGCATTTGCTTGCAGTCCCTACCGCGGGGAATCTTGCCTGTTT 337
|||||
QY 265 CTGGCTGGGGTCTGCTGGGCAACGGTG 291
|||||
Db 338 CTGGCTGGGGTCTGCTGGGCAACGGTG 364
|||||
```

```
RESULT 13
V61253
ID V61253 standard; cDNA; 1119 BP.
AC V61253;
DE 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DEL4.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 27..528
FT /tag= a
FT
PN W09837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillion DC, Xu J;
DR WPI; 98-609886/51.
DR P-PSDB; W71873.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3: Page 108-109; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match 17.6%; Score 257.4; DB 1; Length 1119;
Best Local Similarity 97.8%; Pred. No. 9.6e-48;
Matches 261; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 25 AGTGAATGCAGACTCTTAACCATGCGGCTGGGCTGCACAGTCTTGAAGCCGACCAAG 84
|||||
Db 98 ACTGTTCCAGAACTCTTAACCATGCGGCTGGGCTGCACAGTCTTGAAGCCGACCAAG 157
|||||
QY 85 AGCCAGGAGACCGAGATGGTGGAGGCCCTCTCCGTAAGGAGCCCGAGATACAACAGAC 144
|||||
Db 158 AGCCAGGAGACCGAGATGGTGGAGGCCCTCTCCGTAAGGAGCCCGAGATACAACAGAC 217
|||||
QY 145 CCTTGTCTGCTAAGACCTCATGCTCATCAAGTTGAGAGATCCGTTCCGAGTGTGACA 204
|||||
Db 218 CCTTGTCTGCTAAGACCTCATGCTCATCAAGTTGAGAGATCCGTTCCGAGTGTGACA 277
|||||
QY 205 CCATCCGAGACATCAGCATTTGCTTGCAGTCCCTACCGCGGGGAATCTTGCCTGTTT 264
|||||
Db 278 CCATCCGAGACATCAGCATTTGCTTGCAGTCCCTACCGCGGGGAATCTTGCCTGTTT 337
|||||
QY 265 CTGGCTGGGGTCTGCTGGGCAACGGTG 291
|||||
Db 338 CTGGCTGGGGTCTGCTGGGCAACGGTG 364
|||||

RESULT 14
T18325/C
ID T18325 standard; DNA; 24026 BP.
AC T18325;
DE 05-JUN-1996 (first entry)
DE BRCA1, human breast and ovarian cancer predisposing gene.
KW BRCA1; breast cancer; ovarian cancer; predisposing gene;
KW susceptibility gene; diagnosis; prognosis; gene therapy; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Intron 1..55
FT /tag= a
FT
FT Exon 56..155
FT /tag= b
FT Intron 156..1512
FT /tag= c
FT Exon 1513..1611
```

FT	/*tag- d	1612. .2206	
FT	/*tag- e	/note- "n at 1925-1937 represent an indefinite interval within the Intron"	
FT	/*tag- f	2207. .2260	
FT	/*tag- f	2261. .2677	
FT	/*tag- g	/note- "g at 2569-2581 represent an indefinite interval within the Intron"	
FT	/*tag- h	2678. .2788	
FT	/*tag- h	2789. .3328	
FT	/*tag- i	/note- "n at 3063-3075 represent an indefinite interval within the Intron"	
FT	/*tag- j	3329. .3406	
FT	/*tag- j	3407. .3813	
FT	/*tag- k	/note- "n at 3598-3610 represent an indefinite interval within the Intron"	
FT	/*tag- l	3814. .3902	
FT	/*tag- l	3903. .4224	
FT	/*tag- m	/note- "n at 4076-4088 represent an indefinite interval within the Intron"	
FT	/*tag- n	4225. .4364	
FT	/*tag- n	4365. .6571	
FT	/*tag- o	/note- "o at 4602-4614 represent an indefinite interval within the Intron"	
FT	/*tag- p	6572. .6677	
FT	/*tag- p	6678. .9163	
FT	/*tag- q	9164. .9207	
FT	/*tag- r	9208. .10530	
FT	/*tag- s	10531. .10607	
FT	/*tag- t	10608. .11597	
FT	/*tag- u	/note- "u at 11383-11396 represent an indefinite interval within the Intron"	
FT	/*tag- v	11598. .15023	
FT	/*tag- v	15024. .15424	
FT	/*tag- w	15425. .15511	
FT	/*tag- x	15512. .15952	
FT	/*tag- y	/note- "y at 15647-15659 represent an indefinite interval within the Intron"	
FT	/*tag- z	15953. .16126	
FT	/*tag- z	16127. .16565	
FT	/*tag- aa	/note- "a at 16370-16382 represent an indefinite interval within the Intron"	
FT	/*tag- ab	16566. .16692	
FT	/*tag- ab	16693. .17535	
FT	/*tag- ac	/note- "c at 17290-17302 represent an indefinite interval within the Intron"	
FT	/*tag- ad	17536. .17726	

ET	intron	17727..18416	
ET		/tag- ae	
ET		/note- "n at 18299-18312 represent an indefinite	
ET	exon	interval within the Intron"	
ET		18417..18787	
ET		/tag- af	
ET	intron	18788..19298	
ET		/tag- ag	
ET		/note- "n at 18952-18964 represent an indefinite	
ET	exon	interval within the Intron"	
ET		19299..19386	
ET		/tag- ah	
ET	intron	19387..20190	
ET		/tag- ai	
ET		/note- "n at 19887-19899 represent an indefinite	
ET	exon	interval within the Intron"	
ET		20191..20267	
ET	intron		
ET		/tag- aj	
ET		20268..21094	
ET		/tag- ak	
ET		/note- "n at 20767-20779 represent an indefinite	
ET	exon	interval within the Intron"	
ET		21095..21135	
ET		/tag- al	
ET	intron	21136..21583	
ET		/tag- am	
ET		/note- "n at 21341-21353 represent an indefinite	
ET	exon	interval within the Intron"	
ET		21584..21667	
ET	intron		
ET		/tag- an	
ET		21668..22233	
ET		/tag- ao	
ET		/note- "n at 21921-21933 represent an indefinite	
ET	exon	interval within the Intron"	
ET		22234..22288	
ET	intron		
ET		/tag- ap	
ET		22289..22832	
ET		/tag- aq	
ET		/note- "n at 22567-22579 represent an indefinite	
ET	exon	interval within the Intron"	
ET		22833..22906	
ET		/tag- ar	
ET	intron	22907..23287	
ET		/tag- as	
ET		/note- "n at 23050-23062 represent an indefinite	
ET	exon	interval within the Intron"	
ET		23288..23348	
ET		/tag- at	
ET	intron	23349..23698	
ET		/tag- au	
ET		/note- "n at 23580-23592 represent an indefinite	
ET	exon	interval within the Intron"	
ET		23699..24026	
ET		/tag- av	
ET	misc_feature	2725	
ET		/tag- aw	
ET		/note- "polymorphic site"	
ET	misc_feature	3653	
ET		/tag- ax	
ET		/note- "polymorphic site"	
ET	misc_feature	4391	
ET		/tag- ay	
ET		/note- "polymorphic site"	
ET	misc_feature	4392	
ET		/tag- az	
ET		/note- "polymorphic site"	
ET	misc_feature	6823	
ET		/tag- ba	
ET		/note- "polymorphic site"	
ET	misc_feature	9106	
ET		/tag- bb	
ET		/note- "polymorphic site"	
ET	misc_feature	9207	

```
FT      /tag- bc
FT      /note- "polymorphic site"
FT      /tag- 9376
FT      /tag- bd
FT      /note- "polymorphic site"
FT      /tag- 11908
FT      /tag- be
FT      /note- "polymorphic site"
FT      /tag- 11994
FT      /note- "polymorphic site"
FT      /tag- 12952
FT      /note- "polymorphic site"
FT      /tag- 13004
FT      /note- "polymorphic site"
FT      /tag- 13004
FT      /note- "polymorphic site"
FT      /tag- 13009
FT      /note- "polymorphic site"
FT      /tag- 13048
FT      /note- "polymorphic site"
FT      /tag- 13048
FT      /note- "polymorphic site"
FT      /tag- 13238
FT      /note- "polymorphic site"
FT      /tag- 13448
FT      /note- "polymorphic site"
FT      /tag- 13539
FT      /note- "polymorphic site"
FT      /tag- 13539
FT      /note- "polymorphic site"
FT      /tag- 13951
FT      /note- "polymorphic site"
FT      /tag- 14041
FT      /note- "polymorphic site"
FT      /tag- 14041
FT      /note- "polymorphic site"
FT      /tag- 14046
FT      /note- "polymorphic site"
```

Query Match 15.5%; Score 226.2; DB 1; Length 24026;

Best Local Similarity 87.0%; Pred. No. 1.3e-40;

Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

```
OY 1162 GCTGGCAGAGTGGCTCATGCTGTAATCCAGCACTTTGGAGGC-GAGCGAGCAGAT 1220
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10357 GCTGGCAGAGTGGCTCATGCTGTAATCCAGCACTTTGGAGGC-TGAGCGGGTGAC 10298

OY 1221 CACTTGAGGTAAAGAGTTCAAGACAGCCTGGCCAAATGGTGAATCCTGTCTACTA 1280
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10297 CACTTGAGGTCAATGATTCAGACAGCCTGGCCAAATGGTGAATCCTGTCTACTA 10238

OY 1281 AAAATACAAAAGTTGACTGATATGTTGGCAGCGCCTTAATCCAGCTACTTGGAGG 1340
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10237 AAAATACAAAAGTTGACTGATATGTTGGCAGCGCCTTAATCCAGCTACTTGGAGG 10178

OY 1341 CTGAGCAGAGAAATGCTTGAATATGAGGAGAGAGTTGAAGTGAAGTGAATCAGAC 1400
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10177 CTGAGCAGAGAAATGCTTGAATATGAGGAGAGAGTTGAAGTGAAGTGAATCAGAC 10118

OY 1401 CACTATATCCAGCTGGGCAACAGAGTAAGACTGTCTCAAAAAAAAAAAAAAAAAA 1459
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10117 CACTATATCCAGCTGGGCAACAGAGTAAGACTGTCTCAAAAAAAAAAAAAAAAAA 10059
```

RESULT 15

T17455/c
ID T17455 standard; cDNA: 24025 BP.

AC T17455;

DT 07-OCT-1996 (first entry)

DE Mutated BRCA1 genomic sequence from sample set MSKCC family 19921.

KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;

KW antibody production; germline alteration; probe; lesion neoplasia; human;

```
KW gene therapy; protein replacement therapy; protein mimetic; BRCA1; ds.
OS Homo sapiens.
FH Key
FH exon
      /tag- a
      /note- "exon 1"
      /tag- b
      /note- "intron 1"
      /tag- 1295
      /note- "intron 1"
      /tag- c
      /note- "known polymorphic site"
      /tag- 1513
      /note- "exon 2"
      /tag- d
      /note- "exon 2"
      /tag- e
      /note- "intron 2"
      /tag- f
      /note- "indefinite interval within intron 2"
      /tag- g
      /note- "known polymorphic site"
      /tag- 2207
      /note- "exon 3"
      /tag- h
      /note- "exon 3"
      /tag- i
      /note- "intron 3"
      /tag- j
      /note- "indefinite interval within intron 3"
      /tag- k
      /note- "exon 4"
      /tag- l
      /note- "known polymorphic site"
      /tag- m
      /note- "intron 4"
      /tag- n
      /note- "indefinite interval within intron 4"
      /tag- o
      /note- "exon 5"
      /tag- p
      /note- "intron 5"
      /tag- q
      /note- "indefinite interval within intron 5"
      /tag- r
      /note- "known polymorphic site"
      /tag- s
      /note- "exon 6"
      /tag- t
      /note- "intron 6"
      /tag- u
      /note- "site of 1 nucleotide deletion"
      /tag- v
      /note- "indefinite interval within intron 6"
      /tag- w
      /note- "exon 7"
      /tag- x
      /note- "indefinite interval within intron 7"
      /tag- y
      /note- "exon 8"
      /tag- z
      /note- "indefinite interval within intron 8"
```



```
FT      /*tag- x
FT      /note- "Intron 7"
FT      misc_feature
FT      4391..4392
FT      /*tag- y
FT      /note- "known polymorphic site"
FT      4602..4614
FT      /*tag- z
FT      /note- "indefinite interval within Intron 7"
FT      misc_feature
FT      6538
FT      /*tag- aa
FT      /note- "known polymorphic site"
FT      6572..6677
FT      exon
FT      /*tag- ab
FT      /note- "exon 8"
FT      6678..9163
FT      /*tag- ac
FT      /note- "intron 8"
FT      6823
FT      /*tag- ad
FT      /note- "known polymorphic site"
FT      9106
FT      misc_feature
FT      /*tag- ae
FT      /note- "known polymorphic site"
FT      9164..9209
FT      exon
FT      /*tag- af
FT      /note- "exon 9"
FT      9207
FT      /*tag- ag
FT      /note- "known polymorphic site"
FT      9210..10530
FT      Intron
FT      /*tag- ah
FT      /note- "intron 9"
FT      9376
FT      /*tag- ai
FT      /note- "known polymorphic site"
FT      10531..10607
FT      exon
FT      /*tag- aj
FT      /note- "exon 10"
FT      10608..11597
FT      Intron
FT      /*tag- ak
FT      /note- "intron 10"
FT      11384..11396
FT      misc_feature
FT      /*tag- al
FT      /note- "indefinite interval within Intron 10"
FT      11598..11598
FT      exon
FT      /*tag- am
FT      /note- "exon 11"
FT      11908
FT      misc_feature
FT      /*tag- an
FT      /note- "known polymorphic site"
FT      11994
FT      misc_feature
FT      /*tag- ao
FT      /note- "known polymorphic site"
FT      12952
FT      misc_feature
FT      /*tag- ap
FT      /note- "known polymorphic site"
FT      13004
FT      misc_feature
FT      /*tag- aq
FT      /note- "known polymorphic site"
FT      13009
FT      misc_feature
FT      /*tag- ar
FT      /note- "known polymorphic site"
FT      13048
FT      misc_feature
FT      /*tag- as
FT      /note- "known polymorphic site"
FT      13238
FT      misc_feature
FT      /*tag- at
FT      /note- "known polymorphic site"
FT      13448
FT      misc_feature
FT      /*tag- au
FT      /note- "known polymorphic site"
FT      13539
FT      misc_feature
FT      /*tag- av
```

```
FT      /note- "known polymorphic site"
FT      13951
FT      misc_feature
FT      /*tag- aw
FT      /note- "known polymorphic site"
FT      14041
FT      misc_feature
FT      /*tag- ax
FT      /note- "known polymorphic site"
FT      14046
FT      misc_feature
FT      /*tag- ay
FT      /note- "known polymorphic site"
FT      14475
FT      misc_feature
FT      /*tag- az
FT      /note- "known polymorphic site"
FT      14874
FT      misc_feature
FT      /*tag- ba
FT      /note- "known polymorphic site"
FT      14891
FT      misc_feature
FT      /*tag- bb
FT      /note- "known polymorphic site"
FT      14966
FT      misc_feature
FT      /*tag- bc
FT      /note- "known polymorphic site"
FT      15024..15424
FT      Intron
FT      /*tag- bd
FT      /note- "intron 11"
FT      15284
FT      misc_feature
FT      /*tag- be
FT      /note- "known polymorphic site"
FT      15425..15511
FT      exon
FT      /*tag- bf
FT      /note- "exon 12"
FT      15512..15952
FT      Intron
FT      /*tag- bg
FT      /note- "intron 12"
FT      15647..15659
FT      misc_feature
FT      /*tag- bh
FT      /note- "indefinite interval within Intron 12"
FT      15953..16126
FT      exon
FT      /*tag- bi
FT      /note- "exon 13"
FT      16077
FT      misc_feature
FT      /*tag- bj
FT      /note- "known polymorphic site"
FT      16127..16565
FT      Intron
FT      /*tag- bk
FT      /note- "intron 13"
FT      16243
FT      misc_feature
FT      /*tag- bl
```

```
Query Match      15.5%; Score 226.2; DB 1; Length 24025;
Best Local Similarity 87.0%; Pred. No. 1.3e-40;
Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 1162 GCTGGCAGGCGCTGCTATGCTGTAATCCAGCACTTGGGAGC-GAGGCAGCAGAT 1220
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10356 GCTGGCAGGCGCTGCTATGCTGTAATCCAGCACTTGGGAGCCTGAGCGGCTGAC 10297

QY 1221 CACTTAGGTAAGAGTTCAAGACACCTGGCCAAATGTAATCTCTGTACTA 1280
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10296 CACTTAGGTAAGAGTTCAAGACACCTGGCCAAATGTAATCTCTGTACTA 10237

QY 1281 AAATACAAAAGTAGTGTGATAGTGGCAGGCGCTGTAATCCAGCACTTGGAGG 1340
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10236 AAATACAAAAGTAGTGTGATAGTGGCAGGCGCTGTAATCCAGCACTTGGAGG 10177

QY 1341 CTGAGCAGGAGTAATCTTGAATATGAGGAGCAGAGTTGAATGAGTCAAC 1400
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 10176 CTGAGCAGGAGTAATCTTGAATATGAGGAGCAGAGTTGAATGAGTCAAC 10117

QY 1401 CACTATCTCAGCTGGGCAACAGAGTACTCTCTCTCAAAAAAAAAAAAAA 1459
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 10116 CTTGTACTCCAGCCTGGCAACAGAGTACTCTCTCAAAAAAAAAAAAAA 10058
```

Wed Sep 29 14:27:25 1999

us-09-030-606-174_1.rng

Page 12

Search completed: September 25, 1999, 09:57:00
Job time: 8071 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 07:29:39 ; Search time 165.01 Seconds
(without alignments)
810.594 Million cell updates/sec

Title: US-09-030-606-174

Perfect score: 1459

Sequence: 1 GGTACGCCGACACTGTTC.....TCAAAAAAAAAAAAAAAAAAAA 1459

Scoring table: IDENTITY_NUC

Searched: 176461 seqs, 45838279 residues

Database: Issued_Patents_NA:**

1: /cgn2_6/ptodata/2/1na/3A_COMB.seq:*
2: /cgn2_6/ptodata/2/1na/3B_COMB.seq:*
3: /cgn2_6/ptodata/2/1na/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/1na/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/1na/PCTUS9.COMB.seq:*
6: /cgn2_6/ptodata/2/1na/Packfiles1.seq:**

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	291.8	20.0	871	3	US-08-744-026-2	Sequence 2, Appl1
2	226.2	15.5	6769	2	US-08-480-784-20	Sequence 20, Appl
3	226.2	15.5	6769	2	US-08-483-553-20	Sequence 20, Appl
4	226.2	15.5	6769	2	US-08-487-002-20	Sequence 20, Appl
5	226.2	15.5	6769	2	US-08-483-554B-20	Sequence 20, Appl
6	226.2	15.5	6769	3	US-08-488-011B-20	Sequence 20, Appl
7	226.2	15.5	6769	3	PCT-US95-10202-20	Sequence 20, Appl
8	226.2	15.5	6769	5	PCT-US95-10203-20	Sequence 20, Appl
9	226.2	15.5	6769	5	PCT-US95-10220-20	Sequence 20, Appl
10	219.2	15.0	7676	3	US-08-451-777A-7	Sequence 7, Appl1
11	219.2	15.0	7676	4	US-08-451-778A-7	Sequence 7, Appl1
12	219.2	15.0	7676	4	US-08-998-208-7	Sequence 7, Appl1
13	219.2	15.0	7676	5	PCT-US95-06743-7	Sequence 7, Appl1
14	219.2	15.0	26764	1	US-08-370-975B-1	Sequence 1, Appl1
15	213.4	14.6	20303	1	US-08-370-975B-6	Sequence 6, Appl1
16	212.2	14.5	4833	5	PCT-US94-08806-28	Sequence 28, Appl
17	212.2	14.5	4833	5	PCT-US95-01829-5	Sequence 5, Appl1
18	212.2	14.5	4833	5	PCT-US95-16626-5	Sequence 5, Appl1
19	211.8	14.5	22481	5	PCT-US95-07201-43	Sequence 43, Appl
20	211.2	14.5	246240	4	US-08-724-394A-20	Sequence 20, Appl
21	211.2	14.5	246240	4	US-08-724-394A-21	Sequence 21, Appl
22	211.2	14.5	246240	4	US-08-724-394A-22	Sequence 22, Appl
23	210.6	14.4	17327	1	US-07-906-871-15	Sequence 15, Appl
24	209.8	14.4	2461	3	US-08-832-883-3	Sequence 3, Appl1
25	209.8	14.4	2461	3	US-08-832-883-3	Sequence 3, Appl1
26	209.8	14.4	1811	3	US-08-848-252-1	Sequence 113, App
27	207.8	14.2	1988	4	US-08-257-963B-11	Sequence 11, Appl
28	207.8	14.2	1988	4	PCT-US95-07201-11	Sequence 11, Appl
29	207.6	14.2	7210	4	US-08-257-963B-10	Sequence 10, Appl
30	207.6	14.2	7210	5	PCT-US95-07201-10	Sequence 10, Appl
31	207.6	14.2	22481	5	PCT-US95-07201-43	Sequence 43, Appl
32	206.8	14.0	282	2	US-08-133-629-8	Sequence 8, Appl1
33	203.6	13.9	2477	2	US-08-429-742-1	Sequence 1, Appl1
34	203.2	13.9	7620	1	US-07-767-135-1	Sequence 1, Appl1
35	203.2	13.9	7620	1	US-07-841-652-1	Sequence 1, Appl1
36	202.4	13.9	246240	4	US-08-724-394A-20	Sequence 20, Appl
37	202.4	13.9	246240	4	US-08-724-394A-21	Sequence 21, Appl

ALIGNMENTS

38	202.4	13.9	246240	4	US-08-724-394A-22	Sequence 22, Appl
39	201.4	13.8	3176	4	US-08-910-733-17	Sequence 17, Appl
40	199.4	13.7	8392	1	US-08-080-255-6	Sequence 6, Appl1
41	199.4	13.7	8392	5	PCT-US93-05857-6	Sequence 6, Appl1
42	198.4	13.6	8174	1	US-07-914-281-5	Sequence 5, Appl1
43	198.4	13.6	8174	1	US-08-393-246-5	Sequence 5, Appl1
44	198.4	13.6	11531	1	US-08-068-945A-1	Sequence 1, Appl1
45	198.4	13.6	3373	1	US-08-273-411-2	Sequence 2, Appl1

RESULT 1

US-08-744-026-2
; Sequence 2, Application US/08744026
; Patent No. 5786148
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goll, Surya K.
; TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
; TITLE OF INVENTION: KALIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/744,026
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0154 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-845-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 871 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: Consensus
; US-08-744-026-2

Query Match 20.0% Score 291.8; DB 3; Length 871;
Best Local Similarity 80.2%; Pred. No. 1.5e-64;
Matches 389; Conservative 0; Mismatches 13; Indels 83; Gaps 1;

QY	25	AGTAGTGCAGAGCTCTACACATGCGGCGCTGCACAGTCTTGAGCCGACCAAG	84
DB	224	ACGTGTCAGAGACTCTACACATGCGGCGCTGCACAGTCTTGAGCCGACCAAG	283
QY	85	AGCCAGGAGCCAGATGTGAGAGCCAGCCTCTCGTAGCCAGCCAGATCAACAGAC	144
DB	284	AGCCAGGAGCCAGATGTGAGAGCCAGCCTCTCGTAGCCAGCCAGATCAACAGAC	343

QY 145 CCTTCGCTGACACGCTCATGCTCATCAAGTTGAGCAATCCGTCGAGTCTGACA 204
DB 344 CCTTCGCTGACACGCTCATGCTCATCAAGTTGAGCAATCCGTCGAGTCTGACA 403
QY 205 CCATCCGAGCATCAAGCATGCTTGGCACTGCCCTACCGGGGGAACCTTCCCTGTTT 264
DB 404 ACATCCGAGCATCAAGCATGCTTGGCACTGCCCTACCGGGGGAACCTTCCCTGTTT 463
QY 265 CTGAGTGGGCTGCTGGCGAAGGCTGAGCTACGAGGTGTGTCTGCCCTCTTCAAGA 324
DB 464 CTGAGTGGGCTGCTGGCGAAGGCTGAGCTACGAGGTGTGTCTGCCCTCTTCAAGA 487
QY 325 GGTCTCTGCCCCAGTGGCGGGGGCTGACCCAGAGCTCTGCCGAGAGATGCTTAC 384
DB 487 -----GGCAGATGCTTAC 500
QY 385 CGTGTGAGTGGCTGAGACGCTGCGGTGTCTGTGANGAGTGTCTGACATTAATGATA 444
DB 501 CGTGTGAGTGGCTGAGACGCTGCGGTGTCTGTGANGAGTGTCTGACATTAATGATA 560
QY 445 CCGCGTGTACACCCAGCATGCTTGGCGCGGAGGAGCAAGACAGAGAGTCTG 504
DB 561 CCGCGTGTACACCCAGCATGCTTGGCGCGGAGGAGCAAGACAGAGAGTCTG 620
QY 505 CAACG 509
DB 621 CAACG 625

RESULT 2

US-08-480-784-20/c
Sequence 20, Application US/08480784
Patent No. 5693473

GENERAL INFORMATION:

APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavligian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESS: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,784
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 6769 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-480-784-20

Query Match 15.5% Score 226.2; DB 2; Length 6769;
Best Local Similarity 87.0%; Pred. NO. 9.5e-48;
Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 1162 GCTGGCAGGCTGCTCATGCTGTAATCCAGCATTGGAGGC-GAGGCAGCAGAT 1220
DB 5743 GCTGGCAGGCTGCTCATGCTGTAATCCAGCATTGGAGGCAGGAGG 5664
QY 1221 CACTTGAGTAAAGATTCAAGACCGCTGGCCAAATGTTGAATCTGTCTACTA 1280
DB 5683 CACTTGAGTAAAGATTCAAGACCGCTGGCCAAATGTTGAATCTGTCTACTA 5624
QY 1281 AAATACAAAGTTAGCTGATGTCGAGCGGCGCTGTAATCCAGCTCTGGAGG 1340
DB 5623 AAATACAAAGTTAGCTGATGTCGAGCGGCGCTGTAATCCAGCTCTGGAGG 5564
QY 1341 CTGAGCAGAGAAATTGCTTAATATGGAGCAGAGGTTGAAGTGAATGATCACC 1400
DB 5563 CTGAGCAGAGAAATTGCTTAATATGGAGCAGAGGTTGAAGTGAATGATCACC 5504
QY 1401 CACTTACTCCAGCTGGGCAACAGTAAAGTCTGTCTCAAAAAAAAAAAAAA 1459
DB 5503 CTCTTACTCCAGCTGGGCAACAGTAAAGTCTGTCTCAAAAAAAAAAAAAA 5445

RESULT 3

US-08-483-553-20/c
Sequence 20, Application US/08483553
Patent No. 5709999

GENERAL INFORMATION:

APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavligian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESS: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington

STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,553
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 6769 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-483-553-20

Query Match 15.5%; Score 226.2; DB 2; Length 6769;
Best Local Similarity 87.0%; Pred. No. 9.5e-48;
Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;
QY 1162 GCTGGGACAGGTGCTCCTGTAATCCAGCAGCTTTGGAGGC-GAGCGAGCAGAT 1220
|||||
DB 5743 GCTGGGACAGGTGCTCCTGTAATCCAGCAGCTTTGGAGGCAGCGGGGTGAC 5684
|||||
QY 1221 CACTGAGCTAAGAGTCAAGACAGCGCTGGCCAAATGGGAATCTGCTTACTA 1280
|||||
DB 5683 CACTGAGCTAAGAGTCAAGACAGCGCTGGCCAAATGGGAATCCCACTCTACTA 5624
|||||
QY 1281 AAAATACAAGTTAGCTGATATGTCGAGCGCTGTAATCCAGCTACTTGGAGG 1340
|||||
DB 5623 AAAATACAAGTTAGCTGATATGTCGAGCGCTGTAATCCAGCTACTTGGAGG 5564
|||||
QY 1341 CTGAGGACGAGAAATGCTTGAATATGGAGGACAGAGGTGAAGTGAAGTGAATCAGAC 1400
|||||
DB 5563 CTGAGGACGAGAAATGCTTGAATATGGAGGACAGAGGTGAAGTGAAGTGAATCAGAC 5504
|||||
QY 1401 CACTTATCTCAGCTGGGGCAACAGTAAGCTGTCTCAAAAAAAAAAAAAAAAAAAAA 1459
|||||
DB 5503 CTCTGACTCCAGCTGGGGCAACAGTAAGCTGTCTCAAAAAAAAAAAAAAAAAAAAA 5445
|||||

RESULT 4
US-08-487-002-20/C
Sequence 20, Application US/08487002
Patent No. 5710001
GENERAL INFORMATION:
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Simard, Jacques
APPLICANT: Emi, Mitsuru
APPLICANT: Nakamura, Yusuke
APPLICANT: Dutocher, Francine
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,002
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 6769 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-487-002-20

Query Match 15.5%; Score 226.2; DB 2; Length 6769;
Best Local Similarity 87.0%; Pred. No. 9.5e-48;
Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;
QY 1162 GCTGGGACAGGTGCTCCTGTAATCCAGCAGCTTTGGAGGC-GAGCGAGCAGAT 1220
|||||
DB 5743 GCTGGGACAGGTGCTCCTGTAATCCAGCAGCTTTGGAGGCAGCGGGGTGAC 5684
|||||

```

1 INFORMATION FOR SEQ ID NO: 20:
2 SEQUENCE CHARACTERISTICS:
3 LENGTH: 6769 base pairs
4 TYPE: nucleic acid
5 STRANDEDNESS: double
6 TOPOLOGY: linear
7 MOLECULE TYPE: DNA (genomic)
8 HYPOTHETICAL: NO
9 ANTI-SENSE: NO
10 ORIGINAL SOURCE:
11 ORGANISM: Homo sapiens
12 US-08-483-554B-20
13
14 Query Match 15.5%; Score 226.2; DB 3; Length 6769;
15 Best Local Similarity 87.0%; Pred. No. 9,5e-48;
16 Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;
17
18 QY 1162 GCTGGCGAGGTTGGTGCTCATGCTGTAAATCCAGACACTTGGAGGC-CAGCAGGCAAT 1220
19 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
20 Db 5743 GCTGGCGACGGTGTGCTCAAGCCTGTAAATCCAGACACTTGGAGGCTAGGCGGGTGAC 5684
21
22 QY 1221 CACTTGAGTGAAGAGTTCACAAGACCAGCCTGGCCAAAATGTTGAATCCTGTCTACTA 1280
23 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
24 Db 5663 CACTTGAGTGTATAGTTCACAGACCAGCCTGGCCAACTGGTGAACCCCACTCTACTA 5624
25
26 QY 1261 AAATACAAAAGTTAGCTGATATGTGTGGCAGCGGCCCTGTAAATCCAGACTACTGGGAGG 1340
27 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
28 Db 5623 AAAATACAAAATTATGCGGGTGTGTGGCGCGGCCCTGTAGTCCACACTACTAGAGAG 5564
29
30 QY 1341 CTGAGCGAGGAGAATTGGTTTAATATGGAGGAGCAGAGTTGAAGTAGTTAGATCACAC 1400
31 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
32 Db 5563 CTGAGCGAGGAGTATGAGCTTCAAACCTGGGAGGACAGACCTTGGGAGGAGTAGATCACAC 5504
33
34 QY 1401 CACATTAATCCAGCTGGGGCAACAGAGTAAAGACTCTCTCAAAAAAAAAAAAAA 1459
35 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
36 Db 5503 CTCCTACTCCAGCCTGGGCAACAGAGCAAGCACTCTCTCAAAAAAAAAAAAAA 5445
37
38 RESULT 6
39 US-08-488-011B-20/c
40 Sequence 20, Application US/08488011B
41 Patent No. 5753441
42 GENERAL INFORMATION:
43 APPLICANT: Skolnick, Mark H.
44 APPLICANT: Goldgar, David E.
45 APPLICANT: Mikl, Yoshio
46 APPLICANT: Swenson, Jeff
47 APPLICANT: Kamb, Alexander
48 APPLICANT: Harshman, Keith D.
49 APPLICANT: Shattuck-Eidens, Donna M.
50 APPLICANT: Tavligian, Sean V.
51 APPLICANT: Wiseman, Roger W.
52 APPLICANT: Futreal, P. Andrew
53 TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
54 TITLE OF INVENTION: Susceptibility Gene
55 NUMBER OF SEQUENCES: 85
56 CORRESPONDENCE ADDRESS:
57 ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
58 STREET: 1201 New York Avenue, N.W., Suite 1000
59 CITY: Washington
60 STATE: DC
61 COUNTRY: USA
62 ZIP: 20005
63 COMPUTER READABLE FORM:
64 MEDIUM TYPE: Floppy disk
65 COMPUTER: IBM PC Compatible
66 OPERATING SYSTEM: PC-DOS/MS-DOS
67 SOFTWARE: Patentin Release #1.0, Version #1.30
68 CURRENT APPLICATION DATA:
69 APPLICATION NUMBER: US/08/488,011B
70 FILING DATE: 07-JUN-1995
71 CLASSIFICATION: 435

```

```

PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/409,305
  FILING DATE: 24-MAR-1995
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/348,824
  FILING DATE: 29-NOV-1994
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/308,104
  FILING DATE: 16-SEP-1994
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/300,266
  FILING DATE: 02-SEP-1994
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/289,221
  FILING DATE: 12-AUG-1994
  ATTORNEY/AGENT INFORMATION:
  NAME: Ihnen, Jeffrey L.
  REGISTRATION NUMBER: 28,957
  REFERENCE/DOCKET NUMBER: 24884-109347-09
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: 202-962-4810
  TELEFAX: 202-962-8300
  INFORMATION FOR SEQ ID NO: 20:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 6769 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: double
  TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
  HYPOHETICAL: NO
  ANTI-SENSE: NO
  ORIGINAL SOURCE:
  ORGANISM: Homo sapiens
  US-08-488-011B-20

```

Query Match 15.5%; Score 226.2; DB 3; Length 6769;

Best Local Similarity 87.0%; Pred. No. 9.5e-48; Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

```

OY 1162 GCTGGGAGGCTGCTCATGCTTAATCCAGCACTTGGGAGGC-GAGCGAGCGAGAT 1220
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 5743 GCTGGGAGGCTGCTCATGCTTAATCCAGCACTTGGGAGGC-GAGCGAGCGAGAT 5684
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
OY 1221 CACTTGAGGTAAAGATTCAAGACCGCTGGCCAAATGTTGAATCCTGCTACTA 1280
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 5683 CACTTGAGGTAAAGATTCAAGACCGCTGGCCAAATGTTGAATCCTGCTACTA 5624
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
OY 1281 AAAATACAAAAGTTAGTGTATATGTTGGCAGCGCCCTGTAATCCAGCTACTTGGGAGG 1340
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 5623 AAAATACAAAAGTTAGTGTATATGTTGGCAGCGCCCTGTAATCCAGCTACTTGGGAGG 5564
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
OY 1341 CTGAGCGAGAGAAATGCTGTAATATGGGAGCGAGAGGTTGAAGTGAATGATGATCAGAC 1400
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 5563 CTGAGCGAGAGAAATGCTGTAATATGGGAGCGAGAGGTTGAAGTGAATGATGATCAGAC 5504
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
OY 1401 CACGATCTCCAGCTGGGGAACAGAGTAAGACTCTGCTCAAAAAAATGAAAAAATGAAAAA 1459
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 5503 CTCTGTACTCCAGCTGGGGAACAGAGTAAGACTCTGCTCAAAAAAATGAAAAAATGAAAAA 5445
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||

```

RESULT 7

PCT-US95-10202-20/c

Sequence 20, Application PC/TUS9510202

GENERAL INFORMATION:

APPLICANT: Shattuck-Eidens, Donna M.

APPLICANT: Simard, Jacques

APPLICANT: Emi, Mitsuru

APPLICANT: Nakamura, Yusuke

APPLICANT: Durocher, Francine

TITLE OF INVENTION: In Vivo Mutations and Polymorphisms

TITLE OF INVENTION: In the 17q-Linked Breast and Ovarian Cancer

```

NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10202
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08-308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 6769 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
PCT-US95-10202-20

```

Query Match 15.5%; Score 226.2; DB 5; Length 6769;

Best Local Similarity 87.0%; Pred. No. 9.5e-48; Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

```

OY 1162 GCTGGGAGGCTGCTCATGCTTAATCCAGCACTTGGGAGGC-GAGCGAGCGAGAT 1220
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 5743 GCTGGGAGGCTGCTCATGCTTAATCCAGCACTTGGGAGGC-GAGCGAGCGAGAT 5684
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
OY 1221 CACTTGAGGTAAAGATTCAAGACCGCTGGCCAAATGTTGAATCCTGCTACTA 1280
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 5683 CACTTGAGGTAAAGATTCAAGACCGCTGGCCAAATGTTGAATCCTGCTACTA 5624
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
OY 1281 AAAATACAAAAGTTAGTGTATATGTTGGCAGCGCCCTGTAATCCAGCTACTTGGGAGG 1340
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 5623 AAAATACAAAAGTTAGTGTATATGTTGGCAGCGCCCTGTAATCCAGCTACTTGGGAGG 5564
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
OY 1341 CTGAGCGAGAGAAATGCTGTAATATGGGAGCGAGAGGTTGAAGTGAATGATGATCAGAC 1400
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||

```

DB 5563 CTGAGGACAGAGTATGCTTCAACCGGAGGAGAGTGGAGAGGTGAGATCACAC 5504
QY 1401 CACTATACCTCCAGCTGGGCAACAGAGTACTGTCTCTCAAAAAAAAAAAAAA 1459
DB 5503 CTTCTACTCCAGCTGGGCAACAGAGACTGTCTCAAAAAAAAAAAAAA 5445

RESULT 8

PCT-US95-10203-20/c
Sequence 20, Application PC/TUS9510203
GENERAL INFORMATION:
APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavligian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
TITLE OF INVENTION: Susceptibility Gene
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10203
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824
FILING DATE: 29-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08-308,104
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/300,266
FILING DATE: 02-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,221
FILING DATE: 12-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ihnen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 24884-109347
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4810
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 6769 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
PCT-US95-10203-20

Query Match 15.5%; Score 226.2; DB 5; Length 6769;
Best Local Similarity 87.0%; Pred No. 9.5e-48;
Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 1162 GCTGGGACAGGCTGGCTCATCGCTGTATCCAGCACTTTGGAGGC -GAGGACGAGAGAT 1220
|||||
DB 5743 GCTGGGACAGGCTGGCTCATCGCTGTATCCAGCACTTTGGAGGC -GAGGACGAGAGAT 5684
|||||
QY 1221 CACTGAGGTAGAGGCTTCAACACAGCCTGGCCAAATGTGAAATCCTGTCTACTA 1280
|||||
DB 5683 CACTGAGGTAGAGGCTTCAACACAGCCTGGCCAAATGTGAAATCCTGTCTACTA 5624
|||||
QY 1281 AAAATACAAAGTTAGCTGATATGTTGGGAGGCGCTGTATCCAGCTACTTGGGAGG 1340
|||||
DB 5623 AAAATACAAAGTTAGCTGATATGTTGGGAGGCGCTGTATCCAGCTACTTGGGAGG 5564
|||||
QY 1341 CTGAGGACAGAGATGCTTGAATATGGAGGACAGAGTTGAAGTGAATGATCACAC 1400
|||||
DB 5563 CTGAGGACAGAGTATGCTTCAACCTGGGAGGACAGCTGGGAGAGGTGATGATCACAC 5504
|||||
QY 1401 CACTATACCTCCAGCTGGGCAACAGAGTACTGTCTCTCAAAAAAAAAAAAAA 1459
|||||
DB 5503 CTTCTACTCCAGCTGGGCAACAGAGACTGTCTCTCAAAAAAAAAAAAAA 5445

RESULT 9

PCT-US95-10220-20/c
Sequence 20, Application PC/TUS9510220
GENERAL INFORMATION:
APPLICANT: Skolnick, Mark H.
APPLICANT: Goldgar, David E.
APPLICANT: Miki, Yoshio
APPLICANT: Swenson, Jeff
APPLICANT: Kamb, Alexander
APPLICANT: Harshman, Keith D.
APPLICANT: Shattuck-Eidens, Donna M.
APPLICANT: Tavligian, Sean V.
APPLICANT: Wiseman, Roger W.
APPLICANT: Futreal, P. Andrew
TITLE OF INVENTION: Method for Diagnosing a
TITLE OF INVENTION: Predisposition for Breast and Ovarian Cancer
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10220
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,305
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,824


```

: FILING DATE: 29-NOV-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08-308,104
: FILING DATE: 16-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/300,266
: FILING DATE: 02-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/289,221
: FILING DATE: 12-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Ihnen, Jeffrey L.
: REGISTRATION NUMBER: 28,957
: REFERENCE/DOCKET NUMBER: 24884-109347
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-962-4810
: TELEFAX: 202-962-8300
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6769 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: PCT-US95-10220-20

```

```

Query Match          15.5%; Score 226.2; DB 5; Length 6769;
Best Local Similarity 87.0%; Pred. No. 9.5e-48;
Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

```

```

QY 1162 GCTGGGAGGCGTGCCTGATATCCAGCAGCTTTGGAGGC-GAGGAGGAGAT 1220
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 5743 GCTGGGAGGCGTGCCTGATATCCAGCAGCTTTGGAGGCCTGAGCGGGTGGAC 5684
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
QY 1221 CACTTGAGTGAAGAGCTTCAAGACCCCTGGCCAAATGCTGAATCCTGCTACTA 1280
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 5683 CACTTGAGTGAAGAGCTTCAAGACCCCTGGCCAAATGCTGAATCCTGCTACTA 5624
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
QY 1281 AAATACAAAGTATGCTGATATGTCGAGGCGCCTGTATCCAGTACTTGGAGG 1340
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 5623 AAATACAAAGTATGCTGATATGTCGAGGCGCCTGTATCCAGTACTTGGAGG 5564
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
QY 1341 CTGAGGAGAGAGATTCCTTGAATATGAGGAGGAGGAGTTGAAGTGAATCAGAC 1400
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 5563 CTGAGGAGAGAGATTCCTTGAATATGAGGAGGAGGAGTTGAAGTGAATCAGAC 5504
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
QY 1401 CACTTACTCTCAGCTGGGAGACAGAGTACTCTCTCAAAAAAAAAAAAAA 1459
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 5503 CTCTGACTCTCAGCTGGGAGACAGAGTACTCTCTCAAAAAAAAAAAAAA 5445
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||

```

```

RESULT 10
US-08-451-777A-7/C
: Sequence 7, Application US/08451777A
: Patent No. 5789223
: GENERAL INFORMATION:
: APPLICANT: Bergsma, Derek J.
: APPLICANT: Stambolian, Dwight
: TITLE OF INVENTION: Human Galactokinase Gene
: NUMBER OF SEQUENCES: 33
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SmithKline Beecham Corp./Corporate
: ADDRESSEE: Intellectual Property
: STREET: 709 Swedeland Road/UM2220
: CITY: King of Prussia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19406-0939

```

```

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/451,777A
: FILING DATE: 26-MAY-1995
: CLASSIFICATION: 436
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/10825
: FILING DATE: 23-SEP-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Eagle, Aissa M.
: REGISTRATION NUMBER: 37,126
: REFERENCE/DOCKET NUMBER: P50268-1B
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610-270-5364
: TELEFAX: 610-270-5090
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7676 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-451-777A-7

```

```

Query Match          15.0%; Score 219.2; DB 3; Length 7676;
Best Local Similarity 85.3%; Pred. No. 5.7e-46;
Matches 256; Conservative 0; Mismatches 43; Indels 1; Gaps 1;

```

```

QY 1161 GCTGGGAGGCGTGCCTGATATCCAGCAGCTTTGGAGGC-GAGGAGGAGAT 1219
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 6058 GGTGGGCGGCGGTGCCTACACCTGTATCCAGCAGCTTTGGAGAGCTGAGCGGAG 5999
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
QY 1220 TCACCTGAGTGAAGAGTTCAGACAGCAGCTGGCCAAATGCTGAATCCTGCTACT 1279
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 5998 TTGCTGAGGTCAAGAGCTGAGACCGCTGGCCAAATGCTGAATCCTGCTACT 5939
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
QY 1280 AAATACAAAGTATGCTGATATGTCGAGGCGCCTGTATCCAGTACTTGGAGG 1339
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 5938 AAATACAAAGTATGCTGATATGTCGAGGCGCCTGTATCCAGTACTTGGAGG 5879
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
QY 1340 GCTGAGGAGAGAGATTCCTTGAATATGAGGAGGAGGAGTTGAAGTGAATCAGAC 1399
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 5878 GCTGAGGAGAGAGATTCCTTGAATATGAGGAGGAGGAGTTGAAGTGAATCAGAC 5819
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
QY 1400 CCACTTACTCTCAGCTGGGAGACAGAGTACTCTCTCAAAAAAAAAAAAAA 1459
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||
DB 5818 CCACTTACTCTCAGCTGGGAGACAGAGTACTCTCTCTCAAAAAAAAAAAAAA 5759
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||

```

```

RESULT 11
US-08-451-778A-7/C
: Sequence 7, Application US/08451778A
: Patent No. 5830649
: GENERAL INFORMATION:
: APPLICANT: Bergsma, Derek J.
: APPLICANT: Stambolian, Dwight
: TITLE OF INVENTION: Human Galactokinase Gene
: NUMBER OF SEQUENCES: 33
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SmithKline Beecham Corp./Corporate
: ADDRESSEE: Intellectual Property
: STREET: 709 Swedeland Road/UM2220
: CITY: King of Prussia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19406-0939
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

```

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,208
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/451,777
FILING DATE: 26-MAY-1995
APPLICATION NUMBER: PCT/US94/10825
FILING DATE: 23-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Eagle, Alesia M.
REGISTRATION NUMBER: 37,126
REFERENCE/DOCKET NUMBER: P50268-1B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5364
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 7676 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-998-208-7

```

[illegible]

RESULT 13
PCT-US95-06743-7/C
Sequence 7, Application PC/TUS9506743
GENERAL INFORMATION:
APPLICANT: Bergsma, Derk J.
APPLICANT: Stambolian, Dwight
TITLE OF INVENTION: Human Galactokinase Gene
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESS: Smithkline Beecham Corp./Corporate
ADDRESS: Intellectual Property
STREET: 709 Swedeland Road/0W2220
CITY: King of Prussia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.3

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06743
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: PCT/US94/10825
APPLICATION NUMBER: PCT/US94/10825
FILING DATE: 23-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Sutton, Jeffrey A.
REGISTRATION NUMBER: 34,028
REFERENCE/DOCKET NUMBER: P50268-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5090
TELEFAX: 610-270-5024
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 7676 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US95-06743-7

```

```

Query Match          15.0%; Score 219.2; DB 5; Length 7676;
Best Local Similarity 85.3%; Pred. No. 5.7e-46;
Matches 256; Conservative 0; Mismatches 43; Indels 1; Gaps 1;

```

```

QY 1161 GGCTGGGCGAGGTCGTCATGCTGTATATCCAGACTTTGGGAGGC-GAGGCGAGGCGAGA 1219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6058 GGGTGGGCGGCGGTCGTCACACCTGTATATCCAGACTTTGGGAGGTCGAGGCGAGGCGAGA 5999
QY 1220 TCACCTGAGGTAAGGAGTTCAAGACGACGCGGCAAAATGCGTAATCCTGTGTACT 1279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5998 TTGCGCTAGGTCAGAGACTGAGACGACCTGCGCAACATGGCCAAACCCGCTCTACT 5939
QY 1280 AAAAATCAAAAAGTTAGCTGATATGTTGGCAGGCGGCTGTATATCCAGACTTTGGGAG 1339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5938 AAAAATCAAAAAGTTAGCTGATATGTTGGCAGGCGGCTGTATATCCAGACTTTGGGAG 5879
QY 1340 GCTGAGGCGAGCAAGATTCCTGTAATATGAGGAGGCGAGGTTGAAGTGAAGTGAATCACA 1399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5878 GCTGAGGCGAGCAAGATTCCTGTAATATGAGGAGGCGAGGTTGAAGTGAAGTGAATCACA 5819
QY 1400 CCACTATACCTCAGCTGGGCGAAGAGTAAGACTGTCTCAAAAAAAGAAAAA 1459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5818 CCACTGACCTTCAGCTGGGCGAAGAGTAAGACTGTCTCAAAAAAAGAAAAA 5759

```

```

RESULT 14
US-08-370-975B-1
Sequence 1, Application US/08370975B
Patent No. 5622851
GENERAL INFORMATION:
APPLICANT: Maley, Frank
APPLICANT: Maley, Gladys F.
APPLICANT: Weiner, Karen X.B.
TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/370,975B

```

```

FILING DATE: 10-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Timian, Susan J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 20894/80
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716)263-1636
TELEFAX: (716)263-1600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 26764 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 4q35
US-08-370-975B-1

```

```

Query Match          14.6%; Score 213.4; DB 1; Length 26764;
Best Local Similarity 80.8%; Pred. No. 2.7e-44;
Matches 261; Conservative 0; Mismatches 61; Indels 1; Gaps 1;

```

```

QY 1137 GGAATAATCAAGACTCTACAAAGAGCTGGGCGAGGTCGTCATGCTGTATATCCAGCA 1196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2555 GGTGACAGCCAAAGCCAGAGAGTGAAGCTGGGCGAGTGGCTCACGCTGTATATCCAGCA 2614
QY 1197 CTTTGGAGAG-CGAGGCGAGGAGATCATTGAGTGAAGAGTTCAAGACAGCCTTGCCCA 1255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2615 CTTTGGAGAGTGAAGAGCGGAGGAGATCATTGAGTGAAGAGTTCAAGACAGCCTTGCCCA 2674
QY 1256 AATGCTGAATCTCTGTCTGTATATAAATACAAAGTTAGCTGATATGTTGGCGAGGCG 1315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2675 ACATGGTGAACCCCATCTCTACTTAAATAATTAAGCTGTCCGTTGGTGGCGAGGTCG 2734
QY 1316 CCTGTATCCAGCTACTGTTGGAGGCTGAGGCGAGAGAAATGCTTGAATATGAGGAGGCGA 1375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2735 CCTGTATCTCAGCTACTCTGAGGAGGCTGAGAGAGAGAAATCGTTTAACTCAGAGAGTGG 2794
QY 1376 AGGTTGAAGTGAATGAGATCACACCATATCTCCAGCTGGGCGAAGAGTGAAGTCACTC 1435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2795 AGGTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2854
QY 1436 TGTCTCAAAAAAAGAAAAA 1458
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2855 TGTCTCAAAAAAAGAAAAA 2877

```

```

RESULT 15
US-08-370-975B-6
Sequence 6, Application US/08370975B
Patent No. 5622851
GENERAL INFORMATION:
APPLICANT: Maley, Frank
APPLICANT: Maley, Gladys F.
APPLICANT: Weiner, Karen X.B.
TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/370,975B

```

APPLICATION NUMBER: US/08/370,975B
FILING DATE: 10-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Timian, Susan J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 20894/80
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716)263-1600
TELEFAX: (716)263-1636
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 20303 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 4q35
US-08-370-975B-6

Query Match 14.6%; Score 213.4; DB 1; Length 20303;
Best Local Similarity 80.8%; Pred. No. 2,4e-44;
Matches 261; Conservative 0; Mismatches 61; Indels 1; Gaps 1;
QY 1137 GGAATAATCAAGACTCTACAAAGAGCTGGGAGGCTCATGCGCTTAATCCAGCA 1196
DB 592 GGTCAAGACCAAGCAAGAGTGAAGCTGGGACGATGCTCAGCGCTTAATCCAGCA 651
QY 1197 CTTGGGAGG-CGAGGACGACAGATCCTTGAAGTAAGGATTCAAGACCAGCTGGCCA 1255
DB 652 CTTGGGAGGCTGAGACGGGACATCCTGAGTTCGAGATTTCGAGACCAGCTAGCCA 711
QY 1256 AATGSGTAATCCTGTCTACTATAAATACAAAAGTATAGTGATATGAGGAGCG 1315
DB 712 ACAATGTAACCCCTCTCTACTATAAATACAAAATTAGCTGTGCTGTGGCAGGTG 771
QY 1316 CCTGTAATCCAGTACTTGGAGGCTGAGGACGAGAAATGCTGAATATGAGGAGCGAG 1375
DB 772 CCTGTAATCTAGTACTGCGGAGGCTGAGGAGGAAATCGCTTGAATCAAGAGGTGG 831
QY 1376 AGCTGAAGTACTGATGATCAACACCTATTAATCTCCAGCTGGGGCAACAGAGTAAGACTC 1435
DB 832 AGTTGTACTGAGCCAGACTGCGGCACTGCACCTCCAGCTGGGTGACAGACCAAGACTC 891
QY 1436 TGTCTCAAAAAAAAAAAAAA 1458
DB 892 TGTCTCAAAAAAAAAAAAAA 914

Search completed: September 25, 1999, 07:29:47
Job time: 3424 sec



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPearch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Sep 29 14:01:01 1999; MasPar time 7.55 Seconds
374.490 Million cell updates/sec

Tabular output not generated.

Title: >US-08-931-858E-132
Description: (1-133) from US08931858E.pep
Perfect Score: 997
Sequence: 1 PDARGVPADGEFSSEQVAK.....DRHRMQRPLQLSAAACGGCG 133

Scoring table:
PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

a.geneseqs5
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 30.568; Variance 118.874; scale 0.257

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	645	64.7	134	26	W30067	Mouse persephin.	7.36e-56
2	645	64.7	185	26	W26680	Mature mouse persephi	7.36e-56
3	640	64.2	185	26	W26681	Mature rat persephin.	2.52e-55
4	620	62.2	96	26	W30066	Mature mouse persephi	3.48e-53
5	599	60.1	91	26	W30068	Mature rat persephin.	6.09e-51
6	585	58.7	89	26	W30065	Mature rat persephi	1.90e-49
7	584	58.6	89	26	W30064	Mature mouse persephi	2.43e-49
8	531	53.3	96	26	W30075	Persephin-neurturin c	1.05e-43
9	415	41.6	100	26	W30076	Neurturin-neurturin c	1.77e-31
10	340	34.1	195	25	W13717	Mouse pre-pro-neurtur	1.09e-23
11	338	33.9	197	25	W13716	Human pre-pro-neurtur	1.75e-23
12	326	32.7	100	25	W13711	Mature mouse neurturi	3.00e-22
13	325	32.6	102	25	W13710	Mature human neurturi	3.80e-22
14	244	24.5	104	24	W15744	(Ar93-11e134) trunca	6.08e-14
15	244	24.5	114	24	W15743	(Pro3-Lys37)delAAsn37-	6.08e-14
16	244	24.5	133	14	R79376	Human ATF-2.	6.08e-14

17	244	24.5	134	27	W32106	Human partial glial c	6.08e-14
18	244	24.5	134	25	W23782	Human mature glial ce	6.08e-14
19	244	24.5	134	25	W22027	Glial cell derived ne	6.08e-14
20	244	24.5	134	24	W15706	Glial cell line-deriv	6.08e-14
21	244	24.5	134	25	W18058	Mature human glial ce	6.08e-14
22	244	24.5	134	25	W18052	Mature human glial ce	6.08e-14
23	244	24.5	135	25	W31945	Human glial cell line	6.08e-14
24	244	24.5	135	24	W14930	Human glial cell line	6.08e-14
25	244	24.5	185	14	R79375	Human ATF-1	6.08e-14
26	244	24.5	211	37	W83964	Human glial cell line	6.08e-14
27	243	24.4	160	7	R38298	Human glial derived n	7.65e-14
28	241	24.2	103	24	W15745	(Gly33-11e134) trunca	1.21e-13
29	241	24.2	159	14	R79374	Rat ATF-3.	1.21e-13
30	241	24.2	185	14	R79373	Rat ATF-1.	1.21e-13
31	241	24.2	227	7	R33980	Rat glial-derived neu	1.21e-13
32	228	22.9	133	26	W30069	Human glial-derived n	2.36e-12
33	213	21.4	61	25	W30354	Fragment of human neu	7.06e-11
34	160	16.0	60	25	W30353	Fragment #2 of human	8.61e-06
35	138	13.8	34	25	W30378	Tumour growth inhibit	1.35e-02
36	135	12.5	412	2	R08264	C-terminal monomer of	1.02e-01
37	115	11.5	109	1	P90477	Cattle MIS C-terminal	1.02e-01
38	115	11.5	109	9	R45438	Cattle MIS mature pro	1.02e-01
39	115	11.5	551	14	R76503	Cattle MIS mature pro	1.02e-01
40	115	11.5	575	14	R76500	Sequence of bovine mu	1.02e-01
41	115	11.5	575	2	P70195	Bovine Mullerian inh1	1.02e-01
42	115	11.5	575	1	P90547	C-terminal monomer of	1.25e-01
43	114	11.4	109	1	P90478	Human MIS mature prot	1.25e-01
44	114	11.4	536	14	R76502	Human MIS protein.	1.25e-01
45	114	11.4	560	14	R76501		

ALIGNMENTS

RESULT	1	
ID	W30067 standard; Protein: 134 AA.	
AC	W30067;	
DE	27-MAR-1998 (first entry)	
DE	Mouse persephin.	
KW	Persephin; neurturin; glial-derived neurotrophic factor; GDNF;	
KW	neutrophilic degeneration; haematopoietic cell degeneration;	
KW	cardiac muscle degeneration; amyotrophic lateral sclerosis;	
KW	neuropathy; Alzheimer's disease; Parkinson's disease; stroke;	
KW	Huntington's disease; nervous system tumour; multiple sclerosis;	
KW	osteoporosis; basopenia; lymphopenia; monocytopenia; neutropenia;	
KW	anaemia; thrombocytopenia; stem cell insufficiency; cardiomyopathy;	
KW	congestive heart failure; therapy; mouse.	
OS	Mus musculus.	
PH	Key	Location/Qualifiers
FT	Misc_difference 9	/note="O-glycosylated"
FT	Misc_difference 15	/note="O-glycosylated"
FT	Misc_difference 18	/note="O-glycosylated"
FT	Misc_difference 22	/note="O-glycosylated"
FT	Misc_difference 31	/note="O-glycosylated"
FT	Misc_difference 32	/note="N-glycosylated"
FT	Misc_difference 43	/note="N-glycosylated"
FT	Misc_difference 43	/note="O-glycosylated"
PN	W09733911-AL.	
PD	18-SEP-1997.	
PR	14-MAR-1997; 003461.	
PA	(UNIM) UNIV WASHINGTON.	
PI	Johnson EM, Kotzauer PT, Lampe PA, Milderand JD;	
DR	WPI: 97-470818/43.	
DR	N-PSDB: T90762.	
PT	GDNF-neurturin family related growth factor, Persephin - used to	
PT	prevent or treat cellular, neuronal or non-neuronal, degeneration or	
PT	insufficiency	

PS Claim 3: Page 151: 228PP: English.
CC This polypeptide comprises mouse persephin, a novel growth
CC factor and member of the glial-derived neurotrophic factor-
CC neurturin family. Recombinant persephin can be expressed in
CC host cells utilizing an isolated nucleic acid sequence (see
CC 190762). Persephin polypeptides, and DNA sequences encoding them,
CC can be used in claimed methods to prevent or treat cellular
CC degeneration or insufficiency, such as neuronal degeneration
CC resulting from peripheral neuropathy, amyotrophic lateral
CC sclerosis, Alzheimer's disease, Parkinson's disease, Huntington's
CC disease, ischemic stroke, acute brain or spinal cord injury,
CC nervous system tumours, multiple sclerosis and infection. The
CC cellular degeneration or insufficiency may also comprise
CC haematopoietic cell degeneration or insufficiency resulting from
CC eosinopenia, basopenia, lymphopenia, monocytopenia, neutropenia,
CC anaemia, thrombocytopenia or stem-cell insufficiencies. Also
CC included is cardiac muscle degeneration or insufficiency arising
CC from cardiomyopathy or congestive heart failure. Persephin can be
CC added to a cell culture medium to promote growth and/or
CC differentiation.
CC Sequence 134 AA;

Query Match	64.7%;	Score 645;	DB 26;	length 134;
Best Local Similarity	81.0%;	Pred. No. 7.36e-56;		
Matches	81; Conservative	14; Mismatches	5; Indels	0; Gaps 0;

0y	34	RIRRALSGPQQLSLSLVAELGLGTASDEKVIPTFCMGSCSPRRARKQHGLALRLQGGG	93
Db	35	r1ralasgslwslslpraelqlgyaseekvifvycagssppgarkghlvaiairg	94
0y	94	RANGRCSPRTYDALVGLDRNNKQRLRYQSLAAACGGGG	133
Db	95	rahrgpcqprtsyadvttfiddqhmqqlpqslsaacgsg	134

RESULT	2
ID	W26680 standard; Protein; 185 AA.

DT 27-MAR-1998 (first entry)
DE Mature mouse persephin.
KW Persephin; neuritin; glial-derived neurotrophic factor; GDNF
KW neuronal degeneration; haematopoietic cell degeneration;
KW cardiac muscle degeneration; therapy; mouse.

PN WO9733911-A1.
PD 18-SEP-1997. U03461.
PF 14-MAR-1997; US-615944.
PR 14-MAR-1996; US-615944.
PA (UNIV) ENVI WASHINGTON.
PI Johnson EM, kotzbauer PT, Lampe PA, Milbrandt JD,
DR WPI: 97-470818/43.
N-PSDB: 190802.

p1 GDNF-neurturin family related growth factor, Persephin - used to
 p2 prevent or treat cellular, neuronal or non-neuronal, degeneration or
 p3 insufficiency
 p4
 p5 Example 12: Page 171. 228bp: English.
 p6 This polypeptide sequence was deduced from a mouse persephin gene
 p7 CC (see T90802) obtained from a mouse genomic library. There is an
 p8 CC apparent anomaly in the gene sequence such that the sequence
 p9 CC encoding the RXR cleavage site (see W30081) nd the sequence
 pA CC corresponding to the mature persephin protein (see also W26580) are
 pB CC not co-linear. Instead a second reading frame encodes the cleavage
 pC CC site and mature protein. Persephin is a novel member of the glial-
 pD CC derived neurotrophic factor-neurturin family. Mouse and rat
 pE CC persephin (see W30064-66), and DNA sequences encoding them, are
 pF CC used in claimed methods for preventing or treating neuronal
 pG CC degeneration, haematopoietic cell degeneration and cardiac muscle
 pH CC degeneration or insufficiency.
 pI Sequence 185 AA:

Query Match	64.7%	Score 645;	DB 26;	Length 185;
Best Local Similarity	81.0%;	Pred. No. 7.36e-56;		
Matches ~ 82;	Conservative	14;	Mismatches 5;	Indels 0;
			Gaps 0;	

Db 86 rlpalagscrllwsltlpvaelgigyaseekvlfrycascpcpgeartqhsivlarirgrg 145
 |||:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 34 RLRRALSGPQLWSTLTVSAELGLGYASEEKVIFRYCASCPCPGARTQHGLALARLDGGG 93

```
Db      146 znhgrpcqptsyadvtflddqhhmqqlpqlsaacgcgg 185
      ||||| :|||:|||||:|||||:|||||
QY      94 RAHGPPCRPTRTYDVAFLDDRRRWGRPLQLSAAACGGG 133
```

RESULT	3
ID	W26681 standard; Protein; 185 AA.

DT 27-MAR-1998 (first entry)
DE Mature rat persephin.
KW Persephin; neurtrurin; glia-derived neurotrophic factor; GDNF;
KW neuronal degeneration; haematopoietic cell degeneration;
KW cardiac muscle degeneration; therapy; rat.
KW Rattinge

DR WPI: 57-470818/43.

PT GDNF-neurturin family related growth factor, persephin - used to prevent or treat cellular, neuronal or non-neuronal, degeneration or insufficiency

PT Example 12; Fig 18; 228bp; English.

PS This polypeptide sequence was deduced from a rat persephin gene

CC (see T90805). There is an apparent anomaly in the gene between

CC that encoding the cleavage site for the mature rat perlepin,
CC such that 2 cognate reading frames exist. Perlepin is a novel
CC member of the glial-derived neurotrophic factor-neurturin family.
CC Mouse and rat perlepin polypeptides (see W30064-68), and DNA
CC sequences encoding them, are used in claimed methods for preventing
CC or treating neuronal degeneration, hematopoietic cell degeneration
CC and cardiac muscle degeneration or insufficiency.
SQ Sequence 185 AA;

Query Match	64.28;	Score 640;	DB 26;	Length 185;
Best Local Similarity	80.08;	Pred. No. 2.52e-55;		
Matches	80;	Conservative	13;	Mismatches 7; Indels 0; Gaps 0

D6	86	rLpRLpJcLTlwaSLcLtpvaELgLySseEKLtRrYaaGscpgvqrqthslvarIgrg	145
		:	:
OY	34	RLRRLSPCOLMSTLTLSVALGELGVASEEKVIRYCAGSCPRGARPHQHGALARLGCG	93
		:	:
D6			
OY	94	RAHGGPCCPKPRITVDVAFLDDRNHWQRPLPQLSTAACGGC	133
		:	:
D6	146	rahrppccgptsyadrvflldhhhwqqlpqjsaaacggg	185
		:	:

RESULT	4
ID	W30066 standard; Protein; 96 AA

DT 27-MAR-1998 (first entry)
 DE Mature mouse persephlin.
 KW Persephin; neuritin; glial-derived neurotrophic factor; GDNF;
 KW neuronal degeneration; hematopoietic cell degeneration;
 KW cardiac muscle degeneration; amyotrophic lateral sclerosis;
 KW neuropathy; Alzheimer's disease; Parkinson's disease; stroke;
 KW Huntington's disease; nervous system tumour; multiple sclerosis;
 KW eosinopenia; basopenia; lymphopenia; monocytopenia; neutropenia;
 KW anaemia; thrombocytopenia; stem cell insufficiency; cardiomyopathy;
 KW congestive heart failure; therapy; mouse.

EH	Key	Location/Qualifiers
FT	Misc_difference	10
FT		/note- "O-glycosylated"

FT Misc_difference 12 /note= "O-glycosylated"
FT Misc_difference 24 /note= "O-glycosylated"
FT Misc_difference 36 /note= "O-glycosylated"
FT Misc_difference 36 /note= "O-glycosylated"
FT Misc_difference 43 /note= "O-glycosylated"
FT Misc_difference 46 /note= "O-glycosylated"
FT Misc_difference 67 /note= "O-glycosylated"
FT Misc_difference 68 /note= "O-glycosylated"
FT Misc_difference 73 /note= "O-glycosylated"
FT Misc_difference 88 /note= "O-glycosylated"
FT M09733911-A1.
PD 18-SEP-1997.
PF 14-MAR-1997: U03461.
PR 14-MAR-1986: US-615944.
PL (UNIV) UNIV WASHINGTON.
PI Johnson EM, Kotzbauer PT, Lampe PA, Milbrandt JD;
DR WPI: 97-470818/43.
DR N-PSDB: T90761.
PT GDNF-neurturin family related growth factor. Persephin - used to
PT prevent or treat cellular, neuronal or non-neuronal, degeneration or
PT insufficiency
PS Claim 3; Page 151; 228pp; English.
CC This polypeptide comprises mature mouse persephin, a novel growth
CC factor and member of the glial-derived neurotrophic factor-
CC neurturin family. Recombinant mature persephin can be expressed
CC in host cells utilising an isolated nucleic acid sequence (see
CC T90761). Persephin polypeptides, and DNA sequences encoding them,
CC can be used in claimed methods to prevent or treat cellular
CC degeneration or insufficiency, such as neuronal degeneration
CC resulting from peripheral neuropathy, amyotrophic lateral
CC sclerosis, Alzheimer's disease, Parkinson's disease, Huntington's
CC disease, ischaemic stroke, acute brain or spinal cord injury.
CC nervous system tumours, multiple sclerosis and infection. The
CC cellular degeneration or insufficiency may also comprise
CC haematopoietic cell degeneration or insufficiency resulting from
CC anaemia, basopenia, lymphopenia, monocytopenia, neutropenia,
CC eosinophilia, thrombocytopenia or stem-cell insufficiencies. Also
CC included is cardiac muscle degeneration or insufficiency arising
CC from cardiomyopathy or congestive heart failure. Persephin can be
CC added to a cell culture medium to promote growth and/or
CC differentiation.
Q0 Sequence 96 AA;

Query Match	62.2%	Score 620	DB 26	Length 96
Best Local Similarity	81.3%	Pred. No. 3	48E-53	
Matches	78	Conservative	14	Mismatches 4; Indels 0; Gaps
Dd	1	alagscrlwslilpvaeliglygaseekvifrycagscpgeartqslvialrlgragrh	60	
Oy	38	ALSGCQLMSLTLSTVAELGLGAYASEEKVIFRCAGSCPGARTQHSLALARLQGGGRANG	97	
Dd	61	rpccqptsyadvttfiddqghwqqlpqjsaacgcgg	96	
Oy	98	GPCCRPRTYTDVAFLDDRHRMORLPQLSLAACGCGG	133	

RESULT 5

ID	W30068 standard; Protein: 91 AA.
AC	W30068;
DT	27-MAR-1998 (first entry)
DE	Mature rat persepphin.
KW	Persepphin; neurotrophin; glial-derived neurotrophic factor; GDNF;
KX	neuronal degeneration; haematopoietic cell degeneration;
KY	cardiac muscle degeneration; amyotrophic lateral sclerosis;
KW	neuropathy; Alzheimer's disease; Parkinson's disease; stroke;

KM Huntington's disease; nervous system tumour; multiple sclerosis;
 KM eosinopenia; basopenia; lymphopenia; monocytopenia; neutropenia;
 KM anaemia; thrombocytopenia; stem cell insufficiency; cardiomyopathy;
 KM congestive heart failure; therapy; rat.
 OS Rattus sp.
 PN W09733911-AL.
 PD 18-SEP-1997.
 PF 14-MAR-1997: U03461.
 PR 14-MAR-1996: U03461.
 PA (UNIW) UNIV WASHINGTON.
 PI Johnson EM, Kotzbauer PT, Lampe PA, Milbrandt JD;
 DR WPI: 97-470818/43.
 PT GDNF-neurturin family related growth factor, Persephin - used to
 PT prevent or treat cellular, neuronal or non-neuronal, degeneration or
 PT insufficiency
 PS Claim 3; Page 152; 228pp; English.
 CC This polypeptide comprises mature rat persephin, a novel growth
 CC factor and member of the glial-derived neurotrophic factor-
 CC neurtin family. Recombinant mature persephin can be expressed
 CC in host cells utilizing an isolated nucleic acid sequence.
 CC Persephin polypeptides, and DNA sequences encoding them,
 CC can be used in claimed methods to prevent or treat cellular
 CC degeneration or insufficiency, such as neuronal degeneration
 CC resulting from peripheral neuropathy, amyotrophic lateral
 CC sclerosis, Alzheimer's disease, Parkinson's disease, Huntington's
 CC disease, ischemic stroke, acute brain or spinal cord injury,
 CC nervous system tumours, multiple sclerosis and infection. The
 CC cellular degeneration or insufficiency may also comprise
 CC haematopoietic cell degeneration or insufficiency resulting from
 CC eosinopenia, basopenia, lymphopenia, monocytopenia, neutropenia,
 CC anaemia, thrombocytopenia or stem cell insufficiencies. Also
 CC included is cardiac muscle degeneration or insufficiency arising
 CC from cardiomyopathy or congestive heart failure. Persephin can be
 CC added to a cell culture medium to promote growth and/or
 CC differentiation.
 SQ Sequence 91 AA;
 Query Match 60.1%; Score 599; DB 26; Length 91;
 Best Local Similarity 81.3%; Pred. No. 6,09e-51;
 Matches 74; Conservative 12; Mismatches 5; Indels 0; Gaps 0;
 Db 1 crlslsltlprveidlglyraseekiflfrycascpgvrvtqmslylatirgqgrahyppccg 60
 QY 43 cqlmslsltlsvaelglgyraseekvffrrycascpgvrtqglalrlogqgrahgpcr 102
 Db 61 ptyadvrflldhhbwgqlpqlsaacgcgg 91
 QY 103 pfrtldvaflddrhkrwrlpqlsnaacgcgg 133
 RESULT 6
 ID W30065 standard; Protein: 89 AA.
 AC W30065; 27-MAR-1998 (first entry)
 DT Mature rat persephin.
 DE Persephin; neurturin; glial-derived neurotrophic factor; GDNF;
 KW neuronal degeneration; haematopoietic cell degeneration;
 KW cardiac muscle degeneration; amyotrophic lateral sclerosis;
 KW neuropathy; Alzheimer's disease; Parkinson's disease; stroke;
 KW Huntington's disease; nervous system tumour; multiple sclerosis;
 KW eosinopenia; basopenia; lymphopenia; monocytopenia; neutropenia;
 KW anaemia; thrombocytopenia; stem cell insufficiency; cardiomyopathy;
 KW congestive heart failure; therapy; rat.
 OS Rattus sp.
 PN W09733911-AL.
 PD 18-SEP-1997.
 PF 14-MAR-1997: U03461.
 PR 14-MAR-1996: U03461.
 PA (UNIW) UNIV WASHINGTON.
 PI Johnson EM, Kotzbauer PT, Lampe PA, Milbrandt JD;
 DR WPI: 97-470818/43.
 PT GDNF-neurturin family related growth factor, Persephin - used to
 PT prevent or treat cellular, neuronal or non-neuronal, degeneration or
 PT insufficiency
 PS Claim 3; Page 152; 228pp; English.
 CC This polypeptide comprises mature rat persephin, a novel growth
 CC factor and member of the glial-derived neurotrophic factor-
 CC neurtin family. Recombinant mature persephin can be expressed
 CC in host cells utilizing an isolated nucleic acid sequence.
 CC Persephin polypeptides, and DNA sequences encoding them,
 CC can be used in claimed methods to prevent or treat cellular
 CC degeneration or insufficiency, such as neuronal degeneration
 CC resulting from peripheral neuropathy, amyotrophic lateral
 CC sclerosis, Alzheimer's disease, Parkinson's disease, Huntington's
 CC disease, ischemic stroke, acute brain or spinal cord injury,
 CC nervous system tumours, multiple sclerosis and infection. The
 CC cellular degeneration or insufficiency may also comprise
 CC haematopoietic cell degeneration or insufficiency resulting from
 CC eosinopenia, basopenia, lymphopenia, monocytopenia, neutropenia,
 CC anaemia, thrombocytopenia or stem cell insufficiencies. Also
 CC included is cardiac muscle degeneration or insufficiency arising
 CC from cardiomyopathy or congestive heart failure. Persephin can be
 CC added to a cell culture medium to promote growth and/or
 CC differentiation.
 SQ Sequence 91 AA;
 Query Match 60.1%; Score 599; DB 26; Length 91;
 Best Local Similarity 81.3%; Pred. No. 6,09e-51;
 Matches 74; Conservative 12; Mismatches 5; Indels 0; Gaps 0;
 Db 1 crlslsltlprveidlglyraseekiflfrycascpgvrvtqmslylatirgqgrahyppccg 60
 QY 43 cqlmslsltlsvaelglgyraseekvffrrycascpgvrtqglalrlogqgrahgpcr 102
 Db 61 ptyadvrflldhhbwgqlpqlsaacgcgg 91
 QY 103 pfrtldvaflddrhkrwrlpqlsnaacgcgg 133
 RESULT 6
 ID W30065 standard; Protein: 89 AA.
 AC W30065; 27-MAR-1998 (first entry)
 DT Mature rat persephin.
 DE Persephin; neurturin; glial-derived neurotrophic factor; GDNF;
 KW neuronal degeneration; haematopoietic cell degeneration;
 KW cardiac muscle degeneration; amyotrophic lateral sclerosis;
 KW neuropathy; Alzheimer's disease; Parkinson's disease; stroke;
 KW Huntington's disease; nervous system tumour; multiple sclerosis;
 KW eosinopenia; basopenia; lymphopenia; monocytopenia; neutropenia;
 KW anaemia; thrombocytopenia; stem cell insufficiency; cardiomyopathy;
 KW congestive heart failure; therapy; rat.
 OS Rattus sp.
 PN W09733911-AL.
 PD 18-SEP-1997.
 PF 14-MAR-1997: U03461.
 PR 14-MAR-1996: U03461.
 PA (UNIW) UNIV WASHINGTON.
 PI Johnson EM, Kotzbauer PT, Lampe PA, Milbrandt JD;
 DR WPI: 97-470818/43.
 PT GDNF-neurturin family related growth factor, Persephin - used to
 PT prevent or treat cellular, neuronal or non-neuronal, degeneration or
 PT insufficiency
 PS Claim 3; Page 152; 228pp; English.
 CC This polypeptide comprises mature rat persephin, a novel growth
 CC factor and member of the glial-derived neurotrophic factor-
 CC neurtin family. Recombinant mature persephin can be expressed
 CC in host cells utilizing an isolated nucleic acid sequence.
 CC Persephin polypeptides, and DNA sequences encoding them,
 CC can be used in claimed methods to prevent or treat cellular
 CC degeneration or insufficiency, such as neuronal degeneration
 CC resulting from peripheral neuropathy, amyotrophic lateral
 CC sclerosis, Alzheimer's disease, Parkinson's disease, Huntington's
 CC disease, ischemic stroke, acute brain or spinal cord injury,
 CC nervous system tumours, multiple sclerosis and infection. The
 CC cellular degeneration or insufficiency may also comprise
 CC haematopoietic cell degeneration or insufficiency resulting from
 CC eosinopenia, basopenia, lymphopenia, monocytopenia, neutropenia,
 CC anaemia, thrombocytopenia or stem cell insufficiencies. Also
 CC included is cardiac muscle degeneration or insufficiency arising
 CC from cardiomyopathy or congestive heart failure. Persephin can be
 CC added to a cell culture medium to promote growth and/or
 CC differentiation.
 SQ Sequence 91 AA;
 Query Match 60.1%; Score 599; DB 26; Length 91;
 Best Local Similarity 81.3%; Pred. No. 6,09e-51;
 Matches 74; Conservative 12; Mismatches 5; Indels 0; Gaps 0;
 Db 1 crlslsltlprveidlglyraseekiflfrycascpgvrvtqmslylatirgqgrahyppccg 60
 QY 43 cqlmslsltlsvaelglgyraseekvffrrycascpgvrtqglalrlogqgrahgpcr 102
 Db 61 ptyadvrflldhhbwgqlpqlsaacgcgg 91
 QY 103 pfrtldvaflddrhkrwrlpqlsnaacgcgg 133
 RESULT 6
 ID W30065 standard; Protein: 89 AA.
 AC W30065; 27-MAR-1998 (first entry)
 DT Mature rat persephin.
 DE Persephin; neurturin; glial-derived neurotrophic factor; GDNF;
 KW neuronal degeneration; haematopoietic cell degeneration;
 KW cardiac muscle degeneration; amyotrophic lateral sclerosis;
 KW neuropathy; Alzheimer's disease; Parkinson's disease; stroke;
 KW Huntington's disease; nervous system tumour; multiple sclerosis;
 KW eosinopenia; basopenia; lymphopenia; monocytopenia; neutropenia;
 KW anaemia; thrombocytopenia; stem cell insufficiency; cardiomyopathy;
 KW congestive heart failure; therapy; rat.
 OS Rattus sp.
 PN W09733911-AL.
 PD 18-SEP-1997.
 PF 14-MAR-1997: U03461.
 PR 14-MAR-1996: U03461.
 PA (UNIW) UNIV WASHINGTON.
 PI Johnson EM, Kotzbauer PT, Lampe PA, Milbrandt JD;
 DR WPI: 97-470818/43.
 PT GDNF-neurturin family related growth factor, Persephin - used to
 PT prevent or treat cellular, neuronal or non-neuronal, degeneration or
 PT insufficiency
 PS Claim 3; Page 152; 228pp; English.
 CC This polypeptide comprises mature rat persephin, a novel growth
 CC factor and member of the glial-derived neurotrophic factor-
 CC neurtin family. Recombinant mature persephin can be expressed
 CC in host cells utilizing an isolated nucleic acid sequence.
 CC Persephin polypeptides, and DNA sequences encoding them,
 CC can be used in claimed methods to prevent or treat cellular
 CC degeneration or insufficiency, such as neuronal degeneration
 CC resulting from peripheral neuropathy, amyotrophic lateral
 CC sclerosis, Alzheimer's disease, Parkinson's disease, Huntington's
 CC disease, ischemic stroke, acute brain or spinal cord injury,
 CC nervous system tumours, multiple sclerosis and infection. The
 CC cellular degeneration or insufficiency may also comprise
 CC haematopoietic cell degeneration or insufficiency resulting from
 CC eosinopenia, basopenia, lymphopenia, monocytopenia, neutropenia,
 CC anaemia, thrombocytopenia or stem cell insufficiencies. Also
 CC included is cardiac muscle degeneration or insufficiency arising
 CC from cardiomyopathy or congestive heart failure. Persephin can be
 CC added to a cell culture medium to promote growth and/or
 CC differentiation.
 SQ Sequence 91 AA;
 Query Match 60.1%; Score 599; DB 26; Length 91;
 Best Local Similarity 81.3%; Pred. No. 6,09e-51;
 Matches 74; Conservative 12; Mismatches 5; Indels 0; Gaps 0;
 Db 1 crlslsltlprveidlglyraseekiflfrycascpgvrvtqmslylatirgqgrahyppccg 60
 QY 43 cqlmslsltlsvaelglgyraseekvffrrycascpgvrtqglalrlogqgrahgpcr 102
 Db 61 ptyadvrflldhhbwgqlpqlsaacgcgg 91
 QY 103 pfrtldvaflddrhkrwrlpqlsnaacgcgg 133
 RESULT 6
 ID W30065 standard; Protein: 89 AA.
 AC W30065; 27-MAR-1998 (first entry)
 DT Mature rat persephin.
 DE Persephin; neurturin; glial-derived neurotrophic factor; GDNF;
 KW neuronal degeneration; haematopoietic cell degeneration;
 KW cardiac muscle degeneration; amyotrophic lateral sclerosis;
 KW neuropathy; Alzheimer's disease; Parkinson's disease; stroke;
 KW Huntington's disease; nervous system tumour; multiple sclerosis;
 KW eosinopenia; basopenia; lymphopenia; monocytopenia; neutropenia;
 KW anaemia;

```

Query Match      60.1%  Score 599:  DB 26:  length 91;
Best Local Similarity 81.3%  Pred. No. 6,09e-51;
Matches 74;  Conservative 12;  Mismatches 5;  Indels 0;  Gaps 0;

Db      1  crwsltlplvvelglygaseekilffrycagscpgevrtqslvlarlrgqgrahgrpcq 60
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Oy      43  CQLWSLTLISVAELDLGVAASEKVIFFRCACSCCPRGATIGLALARLQGSGGRHGGCCR 102
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

Db      61  ptsyadvtflddhhhwgqlpqlsaacgcgg 91
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Oy      103  PTRYTDVAFLLDRRHRMORLPOLSPAAACGCCG 133

RESULT      6
ID      W30065 standard; Protein: 89 AA.
AC      W30065;
DT      27-MAR-1998 (first entry)
DE      Mature rat persephin.
KW      Persephin; neuritin; glial-derived neurotrophic factor; GDNF;
KW      neuronal degeneration; haematopoietic cell degeneration;
KW      cardiac muscle degeneration; amyotrophic lateral sclerosis;
KW      neuropathy; Alzheimer's disease; Parkinson's disease; stroke;
KW      Huntington's disease; nervous system tumor; multiple sclerosis;
KW      eosinopenia; basopenia; lymphopenia; monocytopenia; neutropenia;
KW      anaemia; thrombocytopenia; stem cell insufficiency; cardiomyopathy;
KW      congestive heart failure; therapy; rat.
OS      Rattus sp.
PN      W09733911-A1.
PD      18-SEP-1997.
PF      14-MAR-1997:  U034461.
PR      14-MAR-1996:  U5-615944.
PA      (UNIM ) UNIV WASHINGTON.
DR      Johnson EM, Kotzbauer PT, Lampe PA, Milbrandt JD:
      WPL: 97-470818/43.
N-PSD8:  T90760.
PT      GDNF-neurturin family related growth factor, Persephin - used to

```

PN - W02733911-A1.

PA (UNIW) UNIV WASHINGTON.
BT
TOHCOB EV KOTHBVCOB DT 1 2000 PA W11b00004+ TC.

CC cord injury, nervous system tumours, multiple sclerosis and infection;
CC and haematopoietic cell degenerative diseases, e.g. eosinopenia,
CC basopenia, lymphopenia, monocytopenia, neutropenia, anaemia,
CC thrombocytopenia and stem cell insufficiencies. The NT protein and gene
CC are also useful to treat neuroblastomas. Antibodies against NT and
CC oligonucleotides (used as either probes or primers, corresponding to an
CC exon of pre-pro-NT gene or flanking a target sequence) can be used for
CC detecting NT in a sample or detecting mutations in the NT gene. Antisense
CC sequences of the NT gene are used to treat diseases promoted by NT
CC expression e.g. obesity.
SQ Sequence 195 AA;

Query Match 34.1%; Score 340; DB 25; Length 195;
Best Local Similarity 48.2%; Pred. No. 1.09e-23;
Matches 53; Conservative 22; Mismatches 28; Indels 7; Gaps 6;

Db 85 rragprrrararpgarpglrelyvseqlgytsdevlfrycagaceaatlydglr 144
OY 29 HRPRLRLRALSG--PCQLMSLTLSVAELGYSASEKVIYFRYACGSCPRGART-QHGIA 85
DB 145 rlrgrrrrrerarrapccrptaydevsflidshryhtvhsarecac 194
OY 86 -L-ARLQG-GRAHGPGCCRPRTYTD-VAFIDRRHRMORLPOLSAACGC 131

RESULT 11
ID W13716 standard; Protein: 197 AA.

AC W13716;
DT 09-FEB-1998 (first entry)
DE Human pre-pro-neurturin.
KW Neurturin; human; haematopoietic cell; neuronal cell; stem cell; NT gene;
KW neurodegenerative disease; peripheral neuropathy; nervous system tumour;
KW amyotrophic lateral sclerosis; Alzheimer's disease; Parkinson's disease;
KW Huntingdon's disease; ischaemic stroke; acute brain injury; basopenia;
KW acute spinal cord injury; multiple sclerosis; eosinopenia; lymphopenia;
KW monocytopenia; neutropenia; anaemia; thrombocytopenia; neuroblastoma;
KW antibody; obesity; therapy.
OS Homo sapiens.

FT Key Location/Qualifiers
FT Region 1..19 /note= "specifically claimed pre-region"
FT Region 20..95 /note= "specifically claimed pro-region"
FT Peptide 1..95 /note= "pre-pro-peptide"

PD MO9708196-A1.
PF 06-MAR-1997;
PF 27-AUG-1996; U14065.
PR 28-AUG-1995; US-519777.
PA (UNIM) UNIV WASHINGTON.
PI Johnson EM, Kotzbauer PT, Lampe PA, Milbrandt JD;
DR WPI: 97-179176/16.
DR N-PSDB: T61470.
PT A novel growth factor Neurturin - used to treat neuro-degenerative
PT and haematopoietic cell degeneration diseases, e.g. Alzheimer's
PT disease and eosinopenia
PS Claim 13; Page 107-108; 206pp; English.
CC This sequence represents the pre-pro form of human neurturin (NT). NT
CC promotes the growth and differentiation of haematopoietic and neuronal
CC cells, and their stem cells. The NT gene and protein are used to prevent
CC or treat neurodegenerative diseases e.g. peripheral neuropathy,
CC amyotrophic lateral sclerosis, Alzheimer's disease, Parkinson's disease,
CC Huntingdon's disease, ischaemic stroke, acute brain injury, acute spinal
CC cord injury, nervous system tumours, multiple sclerosis and infection;
CC and haematopoietic cell degenerative diseases, e.g. eosinopenia,
CC basopenia, lymphopenia, monocytopenia, neutropenia, anaemia,
CC thrombocytopenia and stem cell insufficiencies. The NT protein and gene
CC are also useful to treat neuroblastomas. Antibodies against NT and
CC oligonucleotides (used as either probes or primers, corresponding to an
CC exon of pre-pro-NT gene or flanking a target sequence) can be used for
CC detecting NT in a sample or detecting mutations in the NT gene. Antisense
CC sequences of the NT gene are used to treat diseases promoted by NT
CC expression e.g. obesity.

SQ Sequence 197 AA;
Query Match 33.9%; Score 338; DB 25; Length 197;
Best Local Similarity 47.3%; Pred. No. 1.75e-23;
Matches 52; Conservative 21; Mismatches 31; Indels 6; Gaps 6;

Db 88 gprrrrarararpgarpglrelyvseqlgytsdevlfrycagaceaatlydglr 146
OY 27 GTHRLRLRALSGPCQLMSLTLSVAELGYSASEKVIYFRYACGSCPRGART-QHGIA 85
DB 147 rlrgrrrrrerarrapccrptaydevsflidshryhtvhsarecac 196
OY 86 -L-ARLQ-GGRAHGPGCCRPRTYTD-VAFIDRRHRMORLPOLSAACGC 131

RESULT 12
ID W13711 standard; Protein: 100 AA.

AC W13711;
DT 11-FEB-1998 (first entry)
DE Mature mouse neurturin.
KW Neurturin; mouse; haematopoietic cell; neuronal cell; stem cell; NT gene;
KW neurodegenerative disease; peripheral neuropathy; nervous system tumour;
KW amyotrophic lateral sclerosis; Alzheimer's disease; Parkinson's disease;
KW Huntingdon's disease; ischaemic stroke; acute brain injury; basopenia;
KW acute spinal cord injury; multiple sclerosis; eosinopenia; lymphopenia;
KW monocytopenia; neutropenia; anaemia; thrombocytopenia; neuroblastoma;
KW antibody; obesity; therapy.
OS Mus musculus.

FT Key Location/Qualifiers
FT Region 6..99 /note= "minimum active fragment"

PD MO9708196-A1.
PF 06-MAR-1997;
PF 27-AUG-1996; U14065.
PR 28-AUG-1995; US-519777.
PA (UNIM) UNIV WASHINGTON.
PI Johnson EM, Kotzbauer PT, Lampe PA, Milbrandt JD;
DR WPI: 97-179176/16.
DR N-PSDB: T61469.

PT A novel growth factor Neurturin - used to treat neuro-degenerative
PT and haematopoietic cell degeneration diseases, e.g. Alzheimer's
PT disease and eosinopenia
PS Disclosure; Page 105; 206pp; English.

CC This sequence represents the mature form of mouse neurturin (NT). NT
CC promotes the growth and differentiation of haematopoietic and neuronal
CC cells, and their stem cells. The NT gene and protein are used to prevent
CC or treat neurodegenerative diseases e.g. peripheral neuropathy,
CC amyotrophic lateral sclerosis, Alzheimer's disease, Parkinson's disease,
CC Huntingdon's disease, ischaemic stroke, acute brain injury, acute spinal
CC cord injury, nervous system tumours, multiple sclerosis and infection;
CC and haematopoietic cell degenerative diseases, e.g. eosinopenia,
CC basopenia, lymphopenia, monocytopenia, neutropenia, anaemia,
CC thrombocytopenia and stem cell insufficiencies. The NT protein and gene
CC are also useful to treat neuroblastomas. Antibodies against NT and
CC oligonucleotides (used as either probes or primers, corresponding to an
CC exon of pre-pro-NT gene or flanking a target sequence) can be used for
CC detecting NT in a sample or detecting mutations in the NT gene. Antisense
CC sequences of the NT gene are used to treat diseases promoted by NT
CC expression e.g. obesity.
SQ Sequence 100 AA;

Query Match

Best Local Similarity 49.5%; Pred. No. 3.00e-22; Length 100;

Matches 47; Conservative 18; Mismatches 25; Indels 5; Gaps 5;

Db 5 pglrelyvseqlgytsdevlfrycagaceaatlydglrllrgrrrerara 64
OY 42 PCQLMSLTLSVAELGYSASEKVIYFRYACGSCPRGART-QHGIA-L-ARLQ-GRAHG 97
DB 65 hpcrptaydevsflidshryhtvhsarecac 99
OY 98 GPCRPRRTYTD-VAFIDRRHRMORLPOLSAACGC 131

OY 34 RLRRALSGPCOLWSLTLSVAELGLGYASEKEKVIFFRYCAGSCPRGART-QHGL-ALARLQG 91
Db 72 lvsdkvsgaccrplafdddlslfddnlvyhllrkhsakrcgc 113
OY 92 QGRAHGG-PCCRPTRYTD-VAFLDDRRHRWQRLPOLSAACGC 131

Search completed: Wed Sep 29 14:01:32 1999
Job time : 31 secs.

(TM)

Distribution rights by Oxford Molecular Ltd

generated

PDARC

1:22810

Existing first 45

Mean 40

and is derived by analysis of the total score distribution.

SUMMARIES

3.45

ALIGNMENTS

七十一

Spontaneously immortalized adult mouse Schwann cells secrete

QY 76 RGAR-TQHGIALARLQGGRAHG-PCCRPRTYD-VAE-LDDRHRM-ORLPOLSAACG 130
 Db 552 C 552
 QY 131 C 131

RESULT 5
 ENTRY WFBOM #type complete
 TITLE Mullerian inhibiting factor precursor - bovine
 ALTERNATE_NAMES Mullerian inhibiting substance (MIS)
 ORGANISM #formal_name Bos primigenius taurus #common_name cattle
 DATE 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 12-Apr-1996
 ACCESSIONS A01398; B01398
 REFERENCE A90879
 #authors Cate, R.L.; Mattaliano, R.J.; Hession, C.; Tizard, R.; Farber, N.M.; Cheung, A.; Ninfa, E.G.; Frey, A.Z.; Gash, D.J.; Chow, E.P.; Fisher, R.A.; Bertoni, J.M.; Torres, G.; Wallner, B.P.; Ramachandran, K.L.; Ragin, R.C.; Mangano, T.F.; MacLaughlin, D.T.; Donahoe, P.K.
 #journal Cell (1986) 45:685-698
 #title Isolation of the bovine and human genes for Mullerian inhibiting substance and expression of the human gene in animal cells.
 #cross-references MWD:86218082
 #accession A01398
 #molecule_type DNA
 #residues 1-14 #label CAL
 #experimental_source newborn calf testis, clones cbmis15 and ps21
 #accession B01398
 #molecule_type mRNA
 #residues 15-575 #label CA2
 COMMENT This glycoprotein, produced by the Sertoli cells of the testis, causes regression of the Mullerian duct. It also is able, in vivo and in vitro, to inhibit the growth of tumors derived from tissues of Mullerian duct origin. Other roles for this protein in gonadal differentiation, meiosis inhibition, and testicle descent are suggested by the low MIS concentrations found in the testis after duct regression and in the adult ovary.
 COMMENT This protein is homologous to the beta transforming growth factor, inhibin alpha chain, and inhibin beta A and B chains. The best area of homology corresponds to the mature protein of all these sequences. All of these proteins are biologically active as disulfide-linked dimers.
 COMMENT Although it does not compete with EGF for receptor binding sites, MIS can inhibit the autophosphorylation of the EGF receptor in vitro.
 CLASSIFICATION #superfamily inhibin
 KEYWORDS cytotxin, glycoprotein, gonadal differentiation, testis
 FEATURE 1-19
 20-24
 25-575
 78,344
 SUMMARY #length 575 #molecular_weight 60623 #checksum 8089

Query Match 11.5%; Score 115; DB 1; Length 575;
 Best Local Similarity 32.8%; Pred. No. 3,85e-05;
 Matches 20; Conservative 17; Mismatches 19; Indels 5; Gaps 5;

Db 514 RNPYGNHVLLKMQARGATLAPCCVPATYATGKLLSLSEERISADHPVNVATECG 573
 QY 76 RGAR-TQHGIALARLQGGRAHG-GPCRPRTYD-VAE-L-DDRHRMORLPOLSAACG 130
 Db 574 C 574
 QY 131 C 131

RESULT 6
 ENTRY S20100 #type complete
 TITLE Mullerian inhibiting factor - mouse
 ORGANISM #formal_name Mus musculus #common_name house mouse
 DATE 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 08-Sep-1997
 ACCESSIONS S20100; S51159
 REFERENCE S20100
 #authors Muensterberg, A.; Lovell-Badge, R.
 #journal Development (1991) 113:613-624
 #title Expression of the mouse anti-Mullerian hormone gene suggests a role in both male and female sexual differentiation.
 #cross-references MWD:92146272
 #accession S20100
 #molecule_type DNA
 #residues 1-555 #label MDE
 #cross-references EMBL:X63240; NID:g49945; PID:g49946
 REFERENCE S51159
 #authors Dresser, D.W.; Hacker, A.; Lovell-Badge, R.; Guertler, D.
 #submission submitted to the EMBL Data Library, January 1995
 #description The genes for anti-Mullerian hormone (AMH) and a spliceosome protein (SAP62) are contiguous.
 #accession S51159
 #status preliminary
 #cross-references EMBL:X63733
 #molecule_type DNA
 #residues 1-41 #label DRE
 #cross-references EMBL:182/3; 219/1; 272/2
 CLASSIFICATION #superfamily inhibin
 SUMMARY #length 555 #molecular_weight 59778 #checksum 3223

Query Match 11.4%; Score 114; DB 2; Length 555;
 Best Local Similarity 34.4%; Pred. No. 5.59e-05;
 Matches 21; Conservative 16; Mismatches 19; Indels 5; Gaps 5;

Db 494 RNPYGNHVLLKMQARGATLAPCCVPATYATGKLLSLSEERISADHPVNVATECG 553
 QY 76 RGAR-TQHGIALARLQGGRAHG-PCCRPRTYD-VAE-L-DDRHRMORLPOLSAACG 130
 Db 554 C 554
 QY 131 C 131

RESULT 7
 ENTRY WFBOM #type complete
 TITLE Mullerian inhibiting factor precursor - human
 ALTERNATE_NAMES anti-Mullerian hormone; mullerian inhibiting substance (MIS)
 ORGANISM #formal_name Homo sapiens #common_name man
 DATE 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 29-May-1998
 ACCESSIONS A01397
 REFERENCE A90879
 #authors Cate, R.L.; Mattaliano, R.J.; Hession, C.; Tizard, R.; Farber, N.M.; Cheung, A.; Ninfa, E.G.; Frey, A.Z.; Gash, D.J.; Chow, E.P.; Fisher, R.A.; Bertoni, J.M.; Torres, G.; Wallner, B.P.; Ramachandran, K.L.; Ragin, R.C.; Mangano, T.F.; MacLaughlin, D.T.; Donahoe, P.K.
 #journal Cell (1986) 45:685-698
 #title Isolation of the bovine and human genes for Mullerian inhibiting substance and expression of the human gene in animal cells.
 #cross-references MWD:86218082
 #accession A01397
 #molecule_type DNA
 #residues 1-560 #label CAT
 #cross-references GB:R03474; NID:g188560; PID:g386953
 COMMENT Although it does not compete with EGF for receptor binding sites, MIS can inhibit the autophosphorylation of the EGF receptor in vitro.
 COMMENT For anti-Mullerian hormone type II receptor, see PIR:JC4335.

gene GDB:AMH
##cross-references GDB:118996; OMIM:261550; OMIM:600957
#map position 19p13.3-19p13.3
#introns 138/1: 185/3: 222/1: 275/2
CLASSIFICATION #superfamily tnfrb
cytokin; glycoprotein; gonadal differentiation; hormone;
KEYWORDS testis

```
FEATURE
1-21      #domain signal sequence #status predicted #label SIG\
22-25     #domain propeptide #status predicted #label PRO\
26-360    #product mullerian inhibiting factor #status predicted
          #label MAT\
64,329    #binding_site carbohydrate (Asn) (covalent) #status
          predicted\
```

492-559	#disulfide_bonds	#status predicted\
525	#disulfide_bonds	interchain #status predicted
SUMMARY	#length 560	#molecular-weight 59192 #checksum 3812

Query Match	11.48;	Score 114;	DB 1;	Length 560;
Best Local Similarity	32.88;	Pred. No. 5.69e-05;		
Matches	20;	Conservative	17;	Mismatches 19; Indels 5; Gaps 5

```
Db      499 RNPFGNVVLLKMGARGAALRPCCVPYTAAGKLISLSEERISAHHYPMVAETCG 558
       | : | : : : | : | | : : | : : | : : | : : |
QY     76 RGAR-TONGTLARLGQGGRANG-GPCRRPTRYD-VAF-L-DDHNRMRQLPQLSAACG 130
```

Db	559 C 559
	1
QY	131 C 131

RESULT	8
ENTRY	S33029
TITLE	hypothetical protein - human herpesvirus 4
ORGANISM	#formal_name human herpesvirus 4, Epstein-Barr virus
DATE	13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997

ACCESSIONS S33029
 REFERENCE S32973
 #authors Farrell, P.J.
 #submission Submitted to the EMBL Data Library, March 1988
 #accession S33029
 #status preliminary
 #molecule_type DNA
 #residues 1-522 ##label FAR
 #cross-references EMBL:V01555; NID:g959074; PID:e25026; PID:g13348888
 #length 522 #molecular-weight 57143 #checksum 8614
 SUMMARY

Query Match	10.48;	Score 104;	DB 2;	Length 522;
Best Local Similarity	32.18;	Pred. No. 2.54e-03;		
Matches	25;	Conservative	19;	Mismatches 28;
			Indels	6;
			Gaps	6;

[illegible]

```

Db      497 WAYADEFILGFVAP TSAH 514
          | : | : | : :
QY      59 YASEKVI FR YCAG SCPR 76

```

RESULT 9

ENTRY	#type complete
A34939	transforming growth factor beta-3 precursor - chicken
TITLE	#formal_name Gallus gallus #common_name chicken
ORGANISM	13-Jul-1990 #sequence_revision 13-Jul-1990 #text_chan
DATE	

ACCESSIONS	A34939; S25850; S36125; S36124; I51181
REFERENCE	A34939

#authors: Jakowlew, S.B.; Dillard, P.J.; Kondalah, P.; Sporn, M.B.;
#journal: Roberts, A.B.
#journal: Mol. Endocrinol. (1988) 2:747-755

#title Complementary deoxyribonucleic acid cloning of a novel transforming growth factor-beta messenger ribonucleic acid from chick embryo chondrocytes.
#cross-references **NCBI:** 89096966
#accession A34939

```
##status      Preliminary; not compared with conceptual translation
##molecule_type  mRNA
##residues      1-412  ##label  JAK
##cross-references  GB:M31154; NID:g212758; PID:g212759
REFERENCE      S25850
#authors      Burt, D.W.; Paton, I.R.; Dey, B.R.
#journal      J. Mol. Endocrinol. (1991) 7:115-183
#title        Comparative analysis of human and chicken transforming growth
               factor-beta-2 and -beta-3 promoters.
#cross-references  W0ID:g2134496
```

```

#accession      S25850
#status         preliminary
#molecule_type DNA
#residues       1-117 #label BUR
#cross-references EMBL:X58127; NID:g65815; PID:g65816
#accession      S36125
#status         preliminary; nucleic acid sequence not shown;

```

```
##molecule_type DNA
##residues 119-172 ##label BU2
##cross_references EMBL:X60055; NID:g396688; PID:g396689
##note the nucleotide sequence was submitted to the EMBL Data
Library, June 1991
#accession S36124
#status preliminary; nucleic acid sequence not shown;
```

##residues 173-322, 'ELPR', 327-412 #label B03
##cross-references EMBL:X6091
##note the nucleotide sequence was submitted to the EMBL Data
Library June 1991
REFERENCE 151181
#authors Jakowlew, S.B.; Teichfelder, R.; Geisler, A.G.; Kim, S.U.;

Santa-Coloma, T.A.; Hubert, J.; Sporn, M.B.; Roberts, A.B.
 Mol. Endocrinol. (1992) 6:1265-1298
 Identification and characterization of the chicken
 transforming growth factor-beta 3 promoter.
 #cross-references MUID:93024487
 #accession I51181
 ##status preliminary
 ##molecule_type DNA
 ##residues 1-117 ##label JAZ
 #cross-references GB:S546000; NID:92571173; PID:92571173

#introns	216/1; 252/1; 309/2; 360/3
#note	list of introns may be incomplete
CLASSIFICATION	#superfamily inhibin

```

FEATURE
1-22      #domain signal sequence #status predicted #label SIG
          glycoprotein; growth factor; growth regulation; homodimer;
          mitogen

```

23-300	#domain	propeptide	#status	predicted	#label	PRO\
301-412	#product	transforming growth factor beta-3	#status	predicted	#label	MAT\
74,142	#binding_site	carbohydrate (Asn)	(covalent)	#status		

SUMMARY #19911412 #molecular-weight 4/103 #checks 2000

Query Match	10.3%	Score 103;	DB 2;	length 412;
Best Local Similarity	29.8%;	Pred. No. 3,66e-03;		
Matches	28;	Mismatches	43;	Indels 5;
				Gaps 4;

37 RALSGPOLMSLTUSV-AELGAGYASEEKYIF-RYCAAGSCP--RGARTOHSLALARIQSG 92

Db 369 NPEASAPCCVPQDLEPTILLYVGRTPKVEQLS 402
|::|||:::|::|||

[illegible]

· Wed Sep 29 14:58:51 1999

US-08-931-858E-132.rpt

Page 7

OY 93 GR-AHGPCCRP 103

Search completed: Wed Sep 29 14:00:44 1999
Job time : 14 secs.

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 12:05:14 ; Search time 1811.29 Seconds
(without alignments)
1588.882 Million cell updates/sec

Title: US-09-030-606-174
Perfect score: 1459
Sequence: 1 GGTCAAGCCGACACGTGTTTC.....TCAAAAAAAAAAAAAAAAAAAA 1459

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

ESR: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: em_est20: *
21: em_est21: *
22: em_est22: *
23: em_est23: *
24: em_est24: *
25: em_est25: *
26: em_est26: *
27: em_est27: *
28: em_est28: *
29: em_est29: *
30: em_est30: *
31: em_est31: *
32: em_est32: *
33: em_est33: *
34: em_est34: *
35: em_est35: *
36: em_est36: *
37: em_est37: *
38: em_est38: *
39: em_est39: *
40: em_est40: *
41: em_est41: *
42: em_est42: *
43: em_est43: *
44: em_est44: *
45: em_est45: *
46: em_est46: *
47: em_est47: *
48: em_est48: *
49: em_est49: *
50: em_est50: *
51: em_est51: *
52: em_est52: *
53: em_est53: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	358.4	24.6	360	50	AI675523
2	286.4	19.6	777	48	AI557281
3	245.4	16.8	569	50	AI686689
4	244.6	16.8	415	35	AA551449
5	242.8	16.6	722	48	AI557025
6	226.2	15.5	481	50	AI669421
7	225.4	15.4	415	48	AI609972
8	223	15.3	407	35	AA535216
9	222.4	15.2	479	26	W96522
10	222.2	15.2	437	36	AA644090
11	220.2	15.1	500	36	AA626040
12	220.2	15.1	454	49	AI634187
13	220	15.1	374	48	AA746911
14	219.8	15.1	434	36	AA630854
15	219.6	15.1	388	39	AA837686
16	219.4	15.0	323	35	AA536040
17	219.2	15.0	484	33	AA410788
18	219.2	15.0	380	36	AA613624
19	218.4	15.0	458	51	AI733856
20	218.2	15.0	408	34	AA486877
21	218	14.9	533	39	AA833875
22	217.2	14.9	470	34	AA456924
23	217.2	14.9	329	36	AA643770
24	217.2	14.9	466	42	AI087040
25	217.2	14.9	365	46	AI419337
26	217.2	14.9	397	46	AI421950
27	217	14.9	415	34	AA515048
28	216.6	14.8	461	30	AA225406
29	216.6	14.8	301	39	AA828592
30	216.4	14.8	528	39	AA833896
31	216.4	14.8	344	41	AI054030
32	216.2	14.8	428	41	AI066646
33	216	14.8	569	53	HS0007223
34	215.2	14.7	474	46	AA601278
35	215.2	14.7	440	36	AA601356
36	215.4	14.8	457	31	AA284247
37	215.4	14.8	480	49	AI635028
38	215.2	14.7	376	29	AA176978
39	215.2	14.7	474	46	AI457313
40	215	14.7	440	36	AA601278
41	214.8	14.7	624	36	AA601356
42	214.8	14.7	623	53	HS0003412
43	214.6	14.7	441	29	AA169245
44	214	14.7	357	45	AI340832
45	214	14.7	413	46	AI333104

ALIGNMENTS

RESULT 1
LOCUS AI675523 360 bp mRNA
DEFINITION WC01F01.x1 NCI_CGAP_Pt28 Homo sapiens cDNA clone IMAGE:2313913 3',
ACCESSION AI675523
NID 94876003
VERSION AI675523.1 GI:4876003

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 360)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 9, 1996 this sequence version replaced gi:1133152.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILMN at:
www.bio.lnlnl.gov/bdrp/image/image.html
Seq primer: -400P from Glibco
High quality sequence stop: 357.
Location/Qualifiers
1..360
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="12"
/clone="IMAGE:2313913"
/clone_1lb="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneids
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 89 a 72 c 69 g 130 t
ORIGIN
Query Match 24.6%; Score 358.4; DB 50; Length 360;
Best Local Similarity 99.7%; Pred. No. 3.4e-57;
Matches 359; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 636 GAGAGAACTGAGAGAAACAGAAATAACACAGAAATAAGAGAGAAAGAGAGAGA 695
DB 360 GAGAGAACTGAGAGAAAGAGAAATAACACAGAAATAAGAGAGAAAGAGAGAGA 301
QY 696 GAAACAGAAACAGACATGGGAGGAGGAGAAACACACACATGAGAGTGAACCTTC 755
DB 300 GAAACAGAAACAGACATGGGAGGAGGAGAAACACACACATGAGAGTGAACCTTC 241
QY 756 CAACAGCATGGGCGCTGAGAGGCGGTGACCTCCACCCCAATAGAAAAATCCTTAATACTT 815
DB 240 CAACAGCATGGGCGCTGAGAGGCGGTGACCTCCACCCCAATAGAAAAATCCTTAATACTT 181
QY 816 TACATCCCAAAACCTGCTGCTGAAATACCTACTGTTAGCGGGAGCCTTACCAATAAC 875
DB 180 TACATCCCAAAACCTGCTGCTGAAATACCTACTGTTAGCGGGAGCCTTACCAATAAC 121
QY 876 ATAAATAGTCGATTAATGATACGTTTATGATCATGATATACCTTGTGGAATTT 935
DB 120 ATAAATAGTCGATTAATGATACGTTTATGATCATGATATACCTTGTGGAATTT 61

QY 936 TTGATATTCTAAGTACACAGATTCGTCGTGAATTTTAAATGTTGCACTTCCT 995
DB 60 TTGATATTCTAAGTACACAGATTCGTCGTGAATTTTAAATGTTGCAATTCCT 1
RESULT 2
LOCUS A1557281 777 bp mRNA EST 23-MAR-1999
DEFINITION PT2.1.15_G12.r tumor2 Homo sapiens CDNA 3', mRNA sequence.
ACCESSION A1557281
NID 94489644
VERSION A1557281.1 GI:4489644
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 777)
AUTHORS Huang,G.M., Ng,W., Farfas,J., Chen,L., Liang,H.A., Gordon,D., Jun
Yu,J. and Hood,L.
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138767.
Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of
Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.
Location/Qualifiers
1..777
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="960H11: 6; 6p21.31-6p22.1; 21q"
/clone_1lb="tumor2"
/note="Organ: prostate; Vector: pBluescript; Directional
CDNA library was constructed using Lambda ZP II Kit
(Stratagene). mRNA was extracted using a frozen prostate
tumor tissue (Mayo Clinics)."
BASE COUNT 162 a 205 c 227 g 157 t 26 others
ORIGIN
Query Match 19.6%; Score 286.4; DB 48; Length 777;
Best Local Similarity 79.8%; Pred. No. 6.1e-44;
Matches 387; Conservative 0; Mismatches 14; Indels 84; Gaps 1;
QY 25 AGTGTGTCAGAGCTCCACACACATCGGGGCTGGGCTGCACAGTCTGAGGCGGACCAAG 84
DB 111 ACTGTTCCAGACTCTCACACCATCGGGGCTGGGCTGCACAGTCTTGAAGCCGACCAAG 170
QY 85 AGCCAGGAGGCAGATGTGTGAGGCGGACGCTCTCTCGTACGGCACCACAGATACACAGAC 144
DB 171 AGCCAGGAGGCAGATGTGTGAGGCGGACGCTCTCTCGTACGGCACCACAGATACACAGAC 230
QY 145 CTTTCTGCTTACAGACTCATGCTCATCAAGTTGAGAGAAATCCGTGTCGAGTCTGACA 204
DB 231 CTTTCTGCTTACAGACTCATGCTCATCAAGTTGAGAGAAATCCGTGTCGAGTCTGACA 290
QY 205 CCATCGGAGAGATGAGATTTGCTTCCGAGTCCCTACCGCGGGGGAACCTTGCCTCGTT 264
DB 291 CCATCGGAGAGATGAGATTTGCTTCCGAGTCCCTACCGCGGGGGAACCTTGCCTCGTT 350
QY 265 CTGAGTGGGGTCTGCTGCGAGACGGTGAAGTCACGGGAGTGTGTCTGCTTCAAGA 324
DB 351 CTGAGTGGGGTCTGCTGCGAGACGG----- 376
QY 325 GGTCTCTGCCCAAGTGGCGGGGGCTGACCCAGAGCTGTGCTCCAGGAGAAATGCCATAC 384
DB 376 -----CAGAAATGCTAC 386

QY 385 CGTGTCTGACGTCGTGAACGTGTCTGTCTGTGANGAGTCTGCANTAGCTCTATGA 444
|||||
Db 387 CGTGTCTGACGTCGTGAACGTGTCTGTCTGTGANGAGTCTGCANTAGCTCTATGA 446
QY 445 CCGCTGTACACCCAGCATGTTCTGTGCGCGGAGGAGGCAAGACCAAGAACTCTCTG 504
|||||
Db 447 CCGCTGTACACCCAGCATGTTCTGTGCGCGGAGGAGGCAAGACCAAGAACTCTCTG 506
QY 505 CCAAC 509
|||||
Db 507 CACG 511

RESULT 3
LOCUS A1686689 569 bp mRNA EST 27-MAY-1999
DEFINITION tu35911.x1 NCI-CGAP_Pt28 Homo sapiens CDNA clone IMAGE:2253092 3'
similar to TR:Q92046 Q92046 PREPROTRYPsin PRECURSOR ;, mRNA
sequence.
ACCESSION A1686689
NID 94897983
VERSION A1686689.1 GI:4897983
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 569)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189584.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmett-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrip/image/image.html

Seq primer: -40UP from GIBCO
High quality sequence stop: 444.

FEATURES
source location/Qualifiers

1.569
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2253092"
/clone_1lb="NCI-CGAP_Pt28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI-CGAP_Pt28 was prepared, and ss
circles were made in vitro. Following BAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneids
985608-986759, 1101192-1101959, and 1217928-1220615)."
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN

125 a 150 c 170 g 124 t
Query Match 16.8%; Score 245.4; DB 50; Length 569;
Best Local Similarity 79.6%; Pred. No. 2,3e-36;

Matches 339; Conservative 0; Mismatches 4; Indels 83; Gaps 1;
QY 84 GAGCCAGGAGCCAGATGTGTGAGAGCCAGCTCTCCGTACAGGACCAGATCAACAGA 143
|||||
Db 569 GAGCCAGGAGCCAGATGTGTGAGAGCCAGCTCTCCGTACAGGACCAGATCAACAGA 510
QY 144 CCGTGTCTGCTTAACGACCTATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTGTAC 203
|||||
Db 509 CCGTGTCTGCTTAACGACCTATGCTCATCAAGTTGGACGAATCCGTGTCCGAGTGTAC 450
QY 204 ACCATCCGAGACATACAGATTGCTGTGCAAGTCCCTACCCGGGGAAGTCTTGCCTGT 253
|||||
Db 449 ACCATCCGAGACATACAGATTGCTGTGCAAGTCCCTACCCGGGGAAGTCTTGCCTGT 390
QY 264 TCTGGCTGGGCTGTGCTGGGAGGAGGAGCTCACGGGTGTGTCTGCCCTTTAAAG 323
|||||
Db 389 TCTGGCTGGGCTGTGCTGGGAGGAGGAGCTCACGGGTGTGTCTGCCCTTTAAAG 365
QY 324 AGGTCTCTGCTCCAGTCCGCGGGGCTGACCCAGAGCTCTGCCAGAGATGCTTA 383
|||||
Db 365 -----GGCAGATGCTTA 353
QY 384 CCGTGTCTGACGTGCGTGAACGTGTGCTGTCTGTGANGAGTCTGCANTAGCTCTATG 443
|||||
Db 352 CCGTGTCTGACGTGCGTGAACGTGTGCTGTCTGTGANGAGTCTGCANTAGCTCTATG 293
QY 444 ACCGCTGTACACCCAGCATGTTCTGTGCGCGGAGGAGGCAAGACCAAGAGACTCT 503
|||||
Db 292 ACCGCTGTACACCCAGCATGTTCTGTGCGCGGAGGAGGCAAGACCAAGAGACTCT 233
QY 504 CCAAC 509
|||||
Db 232 CCAAC 227

RESULT 4
LOCUS AA551449 415 bp mRNA EST 05-SEP-1997
DEFINITION n155605.s1 NCI-CGAP_Pt9 Homo sapiens CDNA clone IMAGE:996416
similar to SW:KLR_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR
; , mRNA sequence.
ACCESSION AA551449
NID 92321701
VERSION AA551449.1 GI:2321701
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 415)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 9, 1995 this sequence version replaced gi:802243.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuagui,
M.D., Michael R. Emmett-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrip/image/image.html

Insert length: 640 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 412.

FEATURES
source

1.415

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:996416"
/clone_1lb="NCI_CGAP_Pr9"
/sex="male"

```

```

/tissue_type="normal prostatic epithelial cells"
/lab_host="DHI0B"
/notes="Organ: prostate; Vector: pAMP10; mRNA made from
normal prostatic epithelial cells. cDNA made by oligo-dT
priming. Non-directionally cloned. Size selected on
agarose gel, average insert size 600 bp. Library made by
K. Kizman, NIH."

```

BASE COUNT	78 a	125 c	129 g	83 t
ORIGIN				

Query Match	16.88;	Score 244.6;	DB 35;	Length 415;
Best Local Similarity	78.28;	Pred. NO. 3.3e-36;		
Matches 344;	Conservative	0;	Mismatches 12;	Indels 84;
				Gaps 1

130 CAGAGTAAACAGACCCCTCTGCTGCTTAAGCAACCTCAAGCTCTAATTAAGCTTGGATGCAATATTCG 188
 QY 70 TTGAGCCCGACCAAGAGCCAGAGGAGCCAGATGCTGGAGGCGACGCTTCCGTATGGGACG 129
 Db 2 TCGAGCCGACCAAGAGCCAGAGGAGCCAGACGCTGGAGGCGACGCTTCCGTATGGGACG 61

D6 62 CAGAGTACACAGACCCCTGGTCGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCG 121

D7 190 TGTCCGAGCTTGACACCATCCGGAGCATCAGCAATTGCTTGGCACTGGCCCTACCGGGGGA 249

QY 250 ACCTTCGCTCGTTTCTTGGGCTGGGGGTCTGCAGCGAAGACTCAGGGGTGTGTC 309
|||||
Db 182 ACTCTGCGCTGTTTCTGGCTGGGGGTCTGCAGCGAACGG----- 222

QY	310	TGCCCTCTTCAAGGAGGCTCTCTCCCAAGTGGCGGGGGCTGACCCAAGCTCTGCGTCC	369
Db	222	-----	222

Db 222 ----CAGAAATGCTACCGTGGCTGCAGTAGTGGGTAAAGCTGTGGTGTCTGTAGAGAGGTCTG. 2777

OY 430 CANTAACTGTATGACCGCGGTGTACCAACCCATGTCGCGCGGCGGAGGAGAGA 489

D0 278 CAGTAAAGCTCCTAATGCCCCCGCTGGTACCCACCCCAATGTTCTTCGCCCGGCGGAGGCACAGA 357

QY 490 CCAGAGAGACTCTCTGTCAACG 509
|||||
DB 338 CCAGAGAGACTCTCTGTCAACG 357
|||||

RESULT	5				
A1557025					
LOCUS	A1557025	722 bp	mRNA	EST	23-MAR-1999

ACCESSION A1557025
NID 94489388
VERSION A1557025.1 GI:4489388
KEYWORDS EST.

ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
REFERENCE	1 (bases 1 to 722)
AUTHORS	Ulanet C, W. No W. Parkes T. Chou I. Fritsch H. A. Gargan D.

TITLE
 Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
 JOURNAL
 Unpublished (1999)
 COMMENT
 On May 18, 1998 this sequence version replaced g1:3138511.

Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of
Washington, Seattle, WA 98195
Tel.: 5106280100
Fax: 5106280108
Email: hungsm@yahoo.com

FEATURES	Location/Qualifiers
source	1. .722

	tumor tissue (Mayo Clinics). "			
BASE COUNT	146 a	156 c	195 g	176 t
ORIGIN				49 others

Query Match	16.6%	Score 242.8;	DB 48;	Length 722;
Best Local Similarity	90.0%;	Pred. No. 6.7e-36;		
Matches 298;	Conservative 0;	Mismatches 25;	Indels 8;	Gaps 4;

OY 96 CAGATGTTGGTGGCCAGGCCTTCCTCGTAGCGGCCAGAGTACAAACAGACCCCTTGCTGCT 155
DB 38 AGCTCTTACACCATTCGGGCTGGGCTCTCACAAGTTTAAGGCCCCAACGAAGACCAGGAGAGC 95
115 AGCTTCTACACCATTCGGGCTGGGCTGTGACAGTCTTTAAGGCCCAACCAAGAAGCCAGGAGAC 174

Db 175 CAGATGGTGGAGGGCAGGCTCTCCGTACGGNACCAGNATGACCAACAGACCCCTTGCTGGCT 234

Q7 156 AACGACCTCATGCTCATCAATTTGGAGGATCCGTGTCCGAGTCTGACACCAATCCGAGC 215

Oy 216 ATCAGCATTTGCTTCGCAGTGGCCCTACCGGGGGAATCTTGCC-TCGTTCCTGGCTGGG 274
|||||
|||
Db 295 ATCAGCATTTGCTTCGNAGTGGCCCTACCGGTGGGAACTTTCCTTTCTGGCTGGG 354
|||||

Oy	275	TCTGC-TGGCCACAGGATAGCTCAGGGTGTGTCTTC-----AAGAGGATCC	329
Dd	355	TCTGTTTGCGAACGGTAGCTCAGGGTGTGTGTTCTGCCCTNTTTAAAGAGATCC	414

```

Db      415  TTTTNCNAGTNTGGGGGGCTNNCCAGAG 445
          | | | | | | | | | | | | | | | |
RESULT  6

```

LOCUS	AI669421	481 bp	mRNA	EST	14-MAY-1999
DEFINITION	ty2a03.01 NCI_CGAP_UF2 Homo sapiens cDNA clone IMAGE:2280773 similar to contains Alu repetitive element; mRNA sequence.				
ACCESSION	AI669421				

VERSION	AT669421.1	GI:4834195
KEYWORDS	Est.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 481)	NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap .
	National Cancer Institute,	Cancer Genome Anatomy Project (CGAP)

JOURNAL
COMMENT

On Dec 20, 1995 this sequence version replaced g1:1130773.

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550

Contact: Robert Strausberg, Ph.D
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuquai,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLT at:
www.bio.llnl.gov/dbfp/image/image.html

FEATURES	Location/Qualifiers
source	1. .407

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="21"
/clone="IMAGE:998337"
/clone_1b="NCI_CGAP_Pri10"
/sex="male"
/tissue_type="Invasive prostate tumor"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pAMP10; mRNA made from
invasive prostate tumor, cDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 600 bp. Library made by D. Krizman,
NIH."

```

Query Match	15.3%	Score 223	DB 35	Length 407
Best Local Similarity	86.3%	Pred. No. 3	2e-32	
Matches 258	Conservative 0	Mismatches 40	Indels 1	Gaps 1
QY	1161	GGCTGGGGCAGGGTGGCTCAATGCCGTAAATCCACGACTTTGGGAGGC	-GAGCGACGCAGA	1219
DB	107	GGCTGGGGCAGTGGGTGCTCAATGCCGTAAATCCACGCACTTTGAGAGGCTAAAGGCAGGCAGA		166
QY	1220	TCACCTGAGAGTAAAGAGATCAAGACACAGCGCTGGCCAAATGGTGAATCCTGTCTGACT		1279
DB	167	TGCGCTGAGTCCAGGGGTTCAAGACCAGCCTGGCCCAACATGTGTAAACCCCGTCTACT		226
QY	1280	AAAAATACAAAAGTTAGCTGATATGTGGGACGCGCTGTAAATCCAGCTACTTGGGAG		1339
DB	227	AAAAATACAAAATTTAGCTGGGCGACGCTGGCAAGTGCCTGTAAATCCAGCTACTTGGGAG		286
QY	1340	GCTGAGGCGAGGAAATGCTTTGAATTTGGGAGGCACAGTGAAGTGAATGATGATCACA		1399
DB	287	GCTGAGGCGAGGAAATGCTTTGAATTTGGGAGGCACAGTGAAGTGAATGATGATGATGTS		346
QY	1400	CCACATACTCCAGCTGGGGCAACAGAGTAAAGACTGTCTCAAAAAAAAAAAAAAAAAAAAA		1458
DB	347	CCACTGCATCCAGCTTGGGTGACAGAGTGAAGACTCCGTCTCAAAAAAAAAAAAAAAAAAAAA		405
RESULT	9	-	3	

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu

This clone is available royalty-free through LNT; contact the
IMAG Consortium (info@imaging.lln.gov) for further information
Insert Length: 3253 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence scope: 457.

FEATURES	Location/Qualifiers
source	1. .479

```

/organism="Homo sapiens"
/db_xref="GD8:1278470"
/db_xref="taxon:9606"
/map="1"
/clone="IMAGE:361767"
/clone_1lb="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: eye; Vector: pTV3D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' GTTACCAATCTGAAGTGGGAGGCGCCGCTTTTCTTTTCTTTTCTTTT 3'] TGTACCAATCTGAAGTGGGAGGCGCCGCTTTTCTTTTCTTTTCTTTT 3' double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pTV3 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)+ RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Fatima Bonaldo."

```

BASE COUNT 89 a 134 c 100 g 152 t 4 others

ORIGIN

Query Match 15.2%; Score 222.4; DB 26; Length 479;
 Best Local Similarity 83.3%; Pred. No. 4, le-32;
 Matches 264; Conservative 0; Mismatches 52; Indels 1; Gaps 1

Y 1144 TCACACTCTACAAAGAGCGTGGCGAGGCGTTCATGCGCTTGAATCCACGACTTTGGG 1203
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 338 TCAAAAGATAAAGGAGCGGAGCAAGTCCTCATGCTTAATCCACGACTTTGGG 279
 Y 1204 AGGC-GAGGCGAGCATCTTGTAGGTAAAGAGTTCAAGACACGCTGGCCAAATGCT 1262.
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 278 AGCGTAGGTGGCAGATCACCTGAGGTGAGAGTGGAGACCAAGCCTGGCCAACTGTGT 219
 QY 1263 GAAATCCTGTCTGTACTTAAATAAATAAAGTTAGCTGGATATGGTGGCAGCGCTGTAA 1322
 Db 218 GAAACCCCGTCTACTACAAANATACAAAANTTACTGTGGTATGTGTCTCTCTGTAA 159
 QY 1323 TCCAGCTACTTGGAGGCTGAGCAGAGAAATTCCTGAATATGAGGAGCAGAGTTGA 1382
 Db 158 TCCAGCTACTTGGAGGCTGAGCAGAGAAATTCATGATGACCAAGGAGGAGCAGAGTTGC 99
 QY 1383 AGGAGTGTGATATCACCACTATCTCCAGCTGGGCGACAGAGTAAAGTCTGTCTCA 1442
 Db 98 AGGAGCGGAGATCACACACTGCACTCCAGCCTGACACAGAGTAAGTCTGTCTAA 39
 QY 1443 AAAAAAAAAAAAAA 1459
 Db 38 AAAAAAAAAAATAA 22

RESULT 10
 AA644090 437 bp mRNA EST 06-MAR-1998
 LOCUS AA644090/c
 DEFINITION IMAGE:845359 3' similar to contains Alu repetitive element; mRNA sequence.

ACCESSION AA644090
 NID 92569308
 VERSION AA644090.1 GI:2569308
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 437)
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Maita, M.,
 Martin, J., Moore, B., Scheinberg, K., Steptoe, M., Tan, F.,
 Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
 WashU-NCI human EST Project
 Unpublished (1997)
 On Sep 19, 1997 this sequence version replaced gi:1520385.

TITLE JOURNAL
 COMMENT

FEATURES
 source
 1. 437
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="845359"
 /clone_lib="Stratagene lung carcinoma 937218"
 /tissue_type="lung carcinoma"
 /cell_line="NCI-H69"
 /dev_stage="cell line NCI-H69"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: lung; Vector: plasmid SK-; Site: 1;
 EcoRI; Site: 2; XhoI; Cloned unidirectionally. Primer:
 Oligo dt. Small cell carcinoma cell line NCI-H69. Average
 insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
 sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5'
 CTCGAGTTTCTTTTCTTTTCTTTT 3'".
 High quality sequence stop: 416.
 Location/Qualifiers

BASE COUNT 98 a 112 c 78 g 149 t

Query Match 15.2%; Score 222.2; DB 36; Length 437;
 Best Local Similarity 85.5%; Pred. No. 4.5e-32;
 Matches 259; Conservative 0; Mismatches 43; Indels 1; Gaps 1;

QY 1158 AGAGCTGGGCGAGGCTGCTCATGCCCTGTAATCCAGACTTTGGGAGG-CGAGGCGGC 1216
 Db 310 ACAGGCTGGGTACAGGGGCTGCTGTAATCCAGACTTTGGGAGGCGGC 251
 QY 1217 AGATCACTTAGGTAAGAGTTCAGACAGCAGCTGGCCAAATGTAATCTGTGT 1276
 Db 250 AGATCACTTAGATCAGAGATTTGAGACAGCCTGGCCAAATGTAATCTGTGT 191
 QY 1277 ACTAAAAATACAAAAGTTAGCTGATATGCTGGCAGGCGCTGTATCCAGTACTTG 1336
 Db 190 ACTAAAAATACAAAATTTAGCCAGGTGTGGTGGCAGGTGCTGTATCCAGTACTG 131
 QY 1337 GAGGCTGAGCGAGCAGCAATTCCTGAATATGGAGCGCAGGTTGAAGTATGAGATC 1396
 Db 130 GAGGCTGAGCGAGCAGCAATTCCTGAATATGGAGCGGTTGAAGTATGAGATC 71

QY 1397 ACACCACTATCTCCAGCTGGGCGACAGAGTAAGTCTGTCTCAAAAAAAAAA 1456
 Db 70 GCACCACTGCACTCCAGCTGGGTGATAGTAAGTCTGTCTCAAAAAAAAAA 11

QY 1457 AAA 1459
 Db 10 AGA 8

RESULT 11
 AA626040 500 bp mRNA EST 15-OCT-1997
 LOCUS AA626040/c
 DEFINITION IMAGE:1055414 3' similar to contains Alu repetitive element; contains element MER22 repetitive element; mRNA sequence.

ACCESSION AA626040
 NID 92538427
 VERSION AA626040.1 GI:2538427
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 500)
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Maita, M.,
 Martin, J., Moore, B., Scheinberg, K., Steptoe, M., Tan, F.,
 Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
 WashU-NCI human EST Project
 Unpublished (1997)
 On Apr 14, 1993 this sequence version replaced gi:503350.

TITLE JOURNAL
 COMMENT

FEATURES
 source
 1. 500
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="1055414"
 /clone_lib="Soares total_fetus_Nb2HR8_9w"
 /dev_stage="8-9 weeks"
 /lab_host="DH10B"
 /note="Vector: pTZ19-3D-Pac (Pharmacia) with a modified
 polylinker; Site: 1; Not I; Site: 2; Eco RI; 1st strand
 was prepared from mRNA obtained from pooled 8-9 week
 (total) fetus material with a Not I - oligo(dt) primer [5'

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewartson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40m13 fwd. ET from Amersham.
 Location/Qualifiers

TGTTCACATCTGAGTGGAGCGCCGCTTAATTTTTTTTTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p773 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo.
 BASE COUNT 115 a 123 c 107 g 154 t 1 others
 ORIGIN

Query Match 15.1%; Score 220.2; DB 36; Length 500;
 Best Local Similarity 81.2%; Pred. No. 1e-31;
 Matches 255; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1140 AAAATCAACCTCTACAAAGAGGCTGGGCTGCTATGCTTATCCAGACTT 1199
 DB 483 AAGATTAAAGAACTTGCTCCCGCAGCGGTGGTGTCTCTATATCCAGCCTT 424
 QY 1200 TGGGAGCGAGCGAGCAGATCACTTGAAGTCAAGACCGCTGGCCAAAT 1259
 DB 423 TCGGCGGCGAGCGAGCGATCACTTGAAGTCAAGACCGCTGGCCAAAT 364
 QY 1260 GGTGAATCTCTCTCTACTATAAAATGCTGATATGCTGGAGCGCCCTG 1319
 DB 363 GGTGAATCCCGCTCTACTATAAAATGCTGATATGCTGGAGCGCCCTG 304
 QY 1320 TATCCACTACTTGGGAGGCTGAGGAGGATTTGCTGATATGGAGGCGAGGT 1379
 DB 303 TATCCACTACTTGGGAGGCTGAGGAGGATTTGCTGATATGGAGGCGAGGT 244
 QY 1380 TGAAGTGAAGTGAAGTCAACCACTTACTCCAGCTGGGCAACAGATGACTGTC 1439
 DB 243 TCAGAGGAGCGAGCGAGCGCTACTCCAGCTGGGCGAGCGAGGAGCTCCGTC 184
 QY 1440 TCAAAAAAAGAA 1453
 DB 183 TCAAAAAAAGAA 170

RESULT 12
 LOCUS A1634187 454 bp mRNA EST 26-APR-1999
 DEFINITION t552a08.x1 NCI-CGAP Kid8 Homo sapiens cDNA clone IMAGE:2232470 3'
 similar to contains Alu repetitive element; contains element TAR1
 repetitive element; mRNA sequence.
 A1634187
 ACCESSION A1634187
 NID 94685517
 VERSION A1634187.1 GI:4685517
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 454)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 18, 1998 this sequence version replaced gi:3138239.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
 Seq primer: -40UP from Gibco

FEATURES
 Source
 Location/Qualifiers
 1..454
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="5"
 /clone="IMAGE:2232470"
 /clone_lib="NCI CGAP Kid8"
 /tissue_type="renal cell tumor"
 /lab_host="DH10B"
 /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.2 kb. Life Technologies catalog #: 11524-014"

BASE COUNT 118 a 103 c 93 g 140 t
 ORIGIN

Query Match 15.1%; Score 220.2; DB 49; Length 454;
 Best Local Similarity 81.8%; Pred. No. 1e-31;
 Matches 266; Conservative 0; Mismatches 58; Indels 1; Gaps 1;

QY 1130 AGAAACAGAAAAATCAAGACTCTACAAAGAGCTGGGCGAGGCTGCTATGTAAT 1189
 DB 325 AGTAATTTATATATAAAATACAAAAAGAGGCTGGCGTGTGCTACAGCTGTAAT 266
 QY 1190 CCCACACTTTGGGAGG-CGAGGCGAGCGAGATCACTTGAAGTGAAGGATCAAGCCAGC 1248
 DB 265 CCCACACTTTGGGAGGCGGAGGCGAGGATCACTTGAAGTGAAGGATCAAGCCAGC 206
 QY 1249 CTGGCCAAATAGTGAATTCCTGCTGCTACTATAAAATCAAAAGTGAAGTGAATGCTG 1308
 DB 205 CTGGCCAAATAGTGAATTCCTGCTGCTACTATAAAATCAAAAGTGAAGTGAATGCTG 146
 QY 1309 GCAGGCGCTGTATCCAGCTACTTGGGAGGCTGGAGGAGGATTTGCTGAATATG 1368
 DB 145 GCGTGCCTGTATCCAGCTACTTGGGAGGCTGGAGGAGGATTTGCTGAATATG 86
 QY 1369 GAGGCGAGGCTGAAGTGAAGTGAAGTCAACCACTATATCCAGCTGGGCGAGAGGT 1428
 DB 85 GAGGCGAGGCTGAAGTGAAGTGAAGTCAACCACTATATCCAGCTGGGCGAGAGGT 26
 QY 1429 AAGACTGTCTCAAAAAAAGAA 1453
 DB 25 GAGATTCCGTCTCAAAAAAAGAA 1

RESULT 13
 LOCUS AA746911 374 bp mRNA EST 22-JAN-1998
 DEFINITION nx59c10.s1 NCI-CGAP A141 Homo sapiens cDNA clone IMAGE:1267506
 similar to contains Alu repetitive element; mRNA sequence.
 AA746911
 ACCESSION AA746911
 NID 92786869
 VERSION AA746911.1 GI:2786869
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 374)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Dec 12, 1995 this sequence version replaced gi:1119179.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
 Ph.D.
 cDNA Library Preparation: David B. Kitzman, Ph.D.

This Page Blank (uspto)

|||||
235 Cyslysphehnglutrpileglulysrthralalaser 248
seq_name: A_Geneseq_36:W69388

seq_documentation_block:
ID W69388 standard; Protein; 205 AA.
AC W69388;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE13 protein.
KW Prostate tumour specific gene; human; prostate cancer; detection;
therapy.
OS Homo sapiens.
FH Key
FT MISC-difference 127 Location/Qualifiers
FT MISC-difference 204 /note= "unspecified amino acid"
FT MISC-difference 204 /note= "unspecified amino acid"
FT MISC-difference 204 /note= "unspecified amino acid"
PD W09837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998: U03690.
PR 09-FEB-1998: US-904809.
PR 25-FEB-1997: US-806596.
PR 01-AUG-1997: US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
PI WPI: 98-480805/41.
DR N-PSDB: V58647.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Example 1; Page 115-116; 141pp; English.
CC This sequence is encoded by a human prostate tumour specific gene, and
CC can be used in the method of the invention. The method is for detecting
CC prostate cancer comprising contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC this protein sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 205 AA;

alignment_scores:
Quality: 1087.00 Length: 205
Ratio: 5.355 Gaps: 0
Percent Similarity: 99.024 Percent Identity: 99.024

alignment_block:

US-09-030-606-175 x W69388 ..

Align seg 1/1 to: W69388 from: 1 to: 205

```
28 ATGGAACAGAAATGTTCTGCTGGGCTGCTGATCCGAGTGGGT 77
|||||
1 Metcunsnngululeuphecysserglyvalleuvalhispogintprya 17
78 GCTGTGACCGCAGACTGTTTCCAGAACTCTTACACCATCGGGTGGGCC 127
|||||
17 lleuseralalalalalalalalalalalalalalalalalal 34
128 TGCCAGCTGTGAGCGGACCAAGAGCCAGGAGCCAGATGTTGAGAGCC 177
|||||
34 euhsserleuglualalaspnluproglyserglmetvalgluala 50
178 AGCCTCTCCGACGACCCAGAGTACAGACTCTGCTGCTAACA 227
|||||
51 Serleuservalalarghispogintpryaargleuleulealalalal 67
228 CTTGATGCTATCATCAAGTTGAGAGCAATCCGTTCCGAGTCTGACACCATCC 277
|||||
67 pleuhtetleuileylserleuaspngluservalsergluserasphrillea 84
```

```
278 GGACATCAGCAATGCTTCGAGTGCCTACCGGGGGAAGTCTTGCTC 327
|||||
84 rgserrleuserlealalaserglincysprothrhalaglyasnsercysleu 100
328 GTNTCTGCTGGGGGTCTGCTGGCGAAGCGAGATGCTACGCTGTGCA 377
|||||
101 valserlelytrpglyleulealalasnnglyargmeprrothrvalleuht 117
378 CTGCGTAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 427
|||||
117 scysvalasnvalservalalserglu***Valcysserlyleuylza 134
428 ACCGCTGTACCAACCCAGCATGTTCTGCGCCGCGGAGGAGCAAGCCAG 477
|||||
134 sspoleuyltrhalsprosermephecysalaglygllyglinspeln 150
478 AAGACTCTGTGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 527
|||||
151 lylaspsercysasnnglyaspserglylyproleuilecysasnngly 167
528 CTTGACAGGCTCTGTCTCTTTCGAAAAGCCCGTGTGCTGCTGCTGCTG 577
|||||
167 rleugnglyleuvalserpheclylalsalaprocysglylnleuyl 184
578 TGCCAGGTGTCTACCAACCTCTGCAAAATTCAGTGTGATAGAGAAA 627
|||||
184 alproglyvalyltrthasnleucyslysphehnglutrpileglulys 200
628 ACCGCTCCAGNCCAGT 642
|||||
201 Thrvalgln***Ser 205
```

seq_name: A_Geneseq_36:W71872

seq_documentation_block:

ID W71872 standard; Protein; 205 AA.

AC W71872;

DT 06-JAN-1999 (first entry)

DE Protein encoded by prostate tumour clone P703 splice variant DE13.

KW Prostate; cancer; tumour; vaccine; immunogen; clone.

OS Homo sapiens.

FH Key Location/Qualifiers

FT MISC-difference 127 /note= "undefined residue"

FT MISC-difference 204 /note= "undefined residue"

FT MISC-difference 204 /note= "undefined residue"

PD W09837093-A2.

PD 27-AUG-1998.

PF 25-FEB-1998: U03492.

PR 09-FEB-1998: US-020956.

PR 25-FEB-1997: US-806099.

PR 01-AUG-1997: US-904804.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J;

PI WPI: 98-609886/51.

DR Polypeptides comprising immunogenic portions of prostate proteins -

PT used in a vaccine for the treatment of prostate cancer

PS Example 3; Page 107-108; 130pp; English.

CC The present sequence is an immunogenic portion of a prostate tumour

CC protein. The immunogen, or the DNA encoding it, can be used as a

CC vaccine for the treatment of prostate cancer. The immunogen was

CC isolated from a prostate tumour cDNA library obtained by subtracting

CC a prostate tumour cDNA expression library with a normal tissue cDNA

CC library.

SQ Sequence 205 AA;

alignment_scores:
Quality: 1087.00 Length: 205
Ratio: 5.355 Gaps: 0
Percent Similarity: 99.024 Percent Identity: 99.024

alignment_block:

US-09-030-606-175 x W71872 ..

Align seg 1/1 to: W71872 from: 1 to: 205

```

28 ATGGAAAACGAATTTGCTGCTGGCGGCTGCTGCTACCTCCGAGTGGGT 77
|||||
1 MetGluAsnGlnLeuPheCysSerGlyValLeuValHisProGlnTrpVal 17
78 GGTGTAGCGCGGCACTGTTTCCAGAACTCTACCACTCCGCGGCTGGGCG 127
|||||
17 LeuSerAlaAlaHisCysPheGlnSerTrpTrpIleGlyLeuGlyL 34
128 TGCACAGTCTTGAGCGCGCAAGACCAAGAGCCAGGAGCCAGATGTGTAGGCG 177
|||||
34 eHisSerLeuGlnAlaAspGlnGluProGlySerGlnMetValGluAla 50
178 AGCCTCCGTTAGCGGCAACCAAGATCAACAGACTTGTGCTGCTTACCA 227
|||||
51 SerLeuSerValArgHisProGluTrpAsnArgLeuLeuLeuAlaAsnAs 67
228 CCTCATGCTCATCAAGTTGAGCAATCCGTTCCGAGTCCGAGTCCGACATCC 277
|||||
67 PheMetLeuIleLeuLysLeuAspLeuSerValSerGluSerAspTrpIleAs 84
278 GGAGCATCAGCAATGCTGCTGCAAGTGCCTTACCGCGGAGCACTTGTGCTTC 327
|||||
84 rGSerIleSerIleAlaSerGlnCysProThrAlaGlyAsnSerCysLeu 100
328 GTTTCGTGGCTGGGGTCTGCTGGCGAAGCGCAAGATCCCTTACCTGCTGCA 377
|||||
101 ValSerGlyTrpLeuLeuAlaAsnGlyArgMetProThrValLeuHis 117
378 CTGCGTGAAGCTGCTGGTGTCTGAGGANGTCTGACAGTAACCTGTATG 427
|||||
117 scysValAsnValSerValValSerLeu***ValCysSerLysLeuTrpAs 134
428 ACCCGCTGTACACCCAGCATGTTCTGCGCCGCGAGGAGGCAAGACCAAG 477
|||||
134 sProLeuTrpHisProSerMetPheCysAlaGlyGlyGlnAspGln 150
478 AAGGATCTCTGCAACGCTGACTGTGGGGGCGCCCTGATCTGCAACGGGTA 527
|||||
151 LysAspSerCysAsnGlyAspSerGlyGlyProLeuIleCysAsnGlyTr 167
528 CTGTCAGAGGCGCTGTGCTTTCGAAAGACCCGTTGGGCAACTTGGGG 577
|||||
167 rLeuGlnGlyLeuValSerPheGlyLysAlaProCysGlyGlnLeuGlyL 184
578 TGGCAGGTGTCTACACCAACTCTGCAAAATTCAGTAGTAGATAGAGAAA 627
|||||
184 alProGlyValTrpThrAsnLeuCysLysPheThrGluTrpIleGluLys 200
628 ACCGTCACAGNCAGT 642
|||||
201 ThrValGln***Ser 205

```

seq_name: A_Geneseq_36:W69387

seq_documentation_block:

ID W69387 standard: Protein: 159 AA.

AC W69387:

DT 08-DEC-1998 (first entry)

DE Prostate tumour specific gene clone DEL protein.

KW Prostate tumour specific gene; human; prostate cancer; detection;

KM therapy.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Misc_difference 103 /note= "unspecified amino acid"

FT Misc_difference 105 /note= "unspecified amino acid"

FT W09837/418-A2.

PD 27-AUG-1998.

```

PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-480805/41.
DR N-PSDB: V58644.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Example 1; Page 112-113; 141pp; English.
CC This sequence is encoded by a human prostate tumour specific gene, and
CC can be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC this protein sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 159 AA;

```

```

alignment_scores:
  Quality: 807.00      Length: 159
  Ratio: 5.206        Gaps: 0
Percent Similarity: 97.484 Percent Identity: 95.597

```

alignment_block:

US-09-030-606-175 x W69387 ..

Align seg 1/1 to: W69387 from: 1 to: 159

```

166 ATGGTGAAGCGCCAGCTCTCCGTACGCGACCCAGATCAACAGACTT 215
|||||
1 MetValGluAlaSerLeuSerValArgHisProGluTrpAsnArgProLe 17
216 GGTGTAGCGCGCACTGATGCTCATCAAGTTGAGCAATCCGTTGGCGAGT 265
|||||
17 uLeuAlaAsnAspLeuMetLeuIleLysLeuAspLeuSerValSerGlu 34
266 CTGACACCATCCGAGAGCATCAGCAATGCTTGCAGTGCCTTACCGCGGG 315
|||||
34 eAspThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGly 50
316 AACCTGTGCTGCTGTTCTGCGTGGGGTCTGCTGGCGAAGCGCAATGCC 365
|||||
51 AsnSerCysLeuValSerGlyTrpLeuLeuAlaAsnGlyArgMetPr 67
366 TACCGGTGCTGCACTGCGTGAAGCTGCGTGTGCTGAGGANGTCTGCA 415
|||||
67 oThrValLeuGlnCysValAsnValSerValValSerGluGluValCys 84
416 GTAAGCTCTATGACCCGCTGTACACCCAGCATGTTCTGCGCGCGCGGA 465
|||||
84 eTrpLeuTrpAspProLeuTrpHisProSerMetPheCysAlaGlyL 100
466 GGGCAAGACCAAGAGACTCTCTGCAACGCTGACTGTGGGGGCGCCCTGAT 515
|||||
101 GlyGln***Gln***AspSerCysAsnGlyAspSerGlyGlyProLeuI 117
516 CTGCAACGGGTACTTGCAGGGCGCTGTGCTTTCGAAAGACCCGCGTGG 565
|||||
117 eCysAsnGlyTrpLeuGlnGlyLeuValSerPheGlyLysAlaProCys 134
566 GCCAATTGGCGTGCAGGTGTCTACACCAACTCTGCAAAATTCAGTAG 615
|||||
134 LysGlnValGlyValProGlyValTrpThrAsnLeuCysLysPheThrGlu 150
616 TGGATAGAGAAAACCGTCCAGNCAGT 642
|||||
151 TrpIleGluLysThrValGlnAlaSer 159

```

seq_name: A_Geneseq_36:W71871

seq_documentation_block:

ID W71871 standard; Protein: 159 AA.

AC W71871;

DT 06-JAN-1999 (first entry)

DE Protein encoded by prostate tumour clone P703 splice variant DEL.

KW Prostate; cancer; tumour; vaccine; immunogen; clone.

OS Homo sapiens.

PH Key location/Qualifiers

FT MISC-difference 103 /note= "undefined residue"

FT MISC-difference 105 /note= "undefined residue"

FT MISC-difference 105 /note= "undefined residue"

PN W09837093-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03492.

PR 09-FEB-1998; US-020956.

PR 25-FEB-1997; US-806099.

PR 01-AUG-1997; US-904804.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J.

DR WPI; 98-609886/51.

PT Polypeptides comprising immunogenic portions of prostate proteins -

PS used in a vaccine for the treatment of prostate cancer

PS Example 3; Page 105; 130pp; English.

CC The present sequence is an immunogenic portion of a prostate tumour

CC protein. The immunogen, or the DNA encoding it, can be used as a

CC vaccine for the treatment of prostate cancer. The immunogen was

CC isolated from a prostate tumour cDNA library obtained by subtracting

CC a prostate tumour cDNA expression library with a normal tissue cDNA

CC library.

SO Sequence 159 AA;

alignment_scores:

Quality: 807.00 Length: 159

Ratio: 5.206 Gaps: 0

Percent Similarity: 97.484 Percent Identity: 95.597

alignment_block:

US-09-030-606-175 x W71871 ..

Align seg 1/1 to: W71871 from: 1 to: 159

```

166 ATGTGTGAGGCCAGCTCTCCGTACGCGCACCCAGATACACAGACTCTT 215
    |||||||
1 MetValGluAlaSerLeuSerValArgHisProGluTyrAsnArgProLeu 17
216 GCTGCTACGACGCTCATGCTCATCAAGTTGACGAATCCGTCCGAGT 265
    |||||||
17 UdenAlaAsnAspLeuMetLeuIleLysLeuAspLysSerValSerGluS 34
266 CTGACACCATCCGAGACATCAGCATTTGCTTCGACAGTCCCTACCGGGG 315
    |||||||
34 eRAspThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGly 50
316 AACTCTTCCTCTGTTCTGGCTGGGGTCTGCTGGGACGACGAGATGCC 365
    |||||||
51 AsnSerCysLeuValSerGlyTrpGlyLeuAlaAsnGlyArgMetArg 67
366 TACCTGTGTGACACTGCTGAGACGTTGCTGGTGTCTGAGAGANGTCTGA 415
    |||||||
67 ThrValLeuGlnCysValAsnValSerValValSerGlnGluValCysS 84
416 GTAACTCTATGACCCGCTGTACACACCCAGCATGTTCTGCGCGGCGGA 465
    |||||||
84 eLysLeuTyrAspProLeuTyrHisProSerMetPheCysAlaGlyGly 100
466 GGGCAGACCCAGAGAGACTCTGTGACAGCGGTGACTGTGGGGGCCCTGAT 515
    |||||||
101 GlyGln***Gln***AspSerCysAsnGlyAspSerGlyGlyProLeuI 117

```

seq_name: A_Geneseq_36:W69389

seq_documentation_block:

ID W69389 standard; Protein: 164 AA.

AC W69389;

DT 08-DEC-1998 (first entry)

DE Prostate tumour specific gene clone DE14 protein.

KW Prostate tumour specific gene; human; prostate cancer; detection;

therapy.

OS Homo sapiens.

PH Key location/Qualifiers

FT MISC-difference 118 /note= "unspecified amino acid"

FT MISC-difference 118 /note= "unspecified amino acid"

PN W09837418-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03690.

PR 09-FEB-1998; US-904809.

PR 25-FEB-1997; US-806596.

PR 01-AUG-1997; US-904809.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J.

DR WPI; 98-480805/41.

DR N-PSDB; V58648.

PT Novel human prostate specific tumour protein and fragments - useful

PT for detecting and treating prostate cancers

PS Example 1; Page 117-118; 141pp; English.

CC This sequence is encoded by a human prostate tumour specific gene, and

CC can be used in the method of the invention. The method is for detecting

CC prostate cancer comprising contacting a biological sample with an agent

CC able to bind an immunogenic portion of a prostate protein (such as

CC this protein sequence). An antibody which binds to an immunogenic

CC portion of the prostate protein, and the method can be used to detect,

CC monitor progression of, or treat prostate cancers. The antibody may

CC also be conjugated to a therapeutic agent for use in therapy of prostate

CC cancers.

SO Sequence 164 AA;

alignment_scores:

Quality: 566.00 Length: 139

Ratio: 4.678 Gaps: 2

Percent Similarity: 87.050 Percent Identity: 82.014

alignment_block:

US-09-030-606-175 x W69389 ..

Align seg 1/1 to: W69389 from: 1 to: 164

```

28 ATGGAAGCAAGATGTCTGCTCGGCGCTCCTGTCGATCCGACAGTGGGT 77
    |||||||
1 MetGluAsnGluLeuPheCysSerGlyValLeuValHisProGlnTrpVal 17
78 GCTGCACGCCACACACTGTTCCAGAACTCCTACACCATCGGGCTGGGCGC 127
    |||||||
17 IleuSerIleAlaHisCysPheGlnAsnSerTyrThrIleGlyLeuGlyL 34
128 TGCACAGTCTTGAGGCCGACCAAGAGCGAGGAGCCAGATGTTGAGAGCC 177
    |||||||
34 euHisSerLeuGluAlaAspGlnGluProGlySerGlnMetValGluAla 50
178 AGCTCTCCGTACGGCACCCAGAGTACACAGACTCTTGCTCGCTACGA 227

```

```

|||||
51 SerLeuSerValAlaArgHisProGluTyrAsnArgProLeuLeuAlaAsnAs 67
228 CCGATCATCTATCAAGTTGGACGAATCCGTCGAGTCTGACACCATCC 277
|||||
67 PluMetLeuIleLysLeuAspGluSerValSerGluSerAspThrIleA 84
278 GGAGCATCAGCATTCGTCGACAGTCCCTACCCGGGGAACTCTTGCCCTC 327
|||||
84 rgserrIleSerIleAlaSerGlnCysProThrAlaGlyAsnSerCysLeu 100
328 GNTCTGGCTGGGGCTGCTGCGGCAACGGCAAGTCCCTACCGTCTGCA 377
|||||
101 ValSerGlyTrpGlyLeuLeuAlaAsnAspAlaVal..... 112
378 CTGCGTGAACGTGCGGTGCTGTGAGGANGTC.....TGCA 415
113 .....IleAlaIleGlnSer***ThrValGlyGlyTrpGluCysG 126
416 GTAAGCTCTATGACCCG 432
126 IuylsLeuSerGlnPro 131
seq_name: A_Geneseq_36:W71873

```

```

seq_documentation_block:
ID W71873 standard; Protein: 164 AA.
AC W71873:
DT 06-JAN-1999 (first entry)
DE Protein encoded by prostate tumour clone P703 splice variant DE14.
KW Prostate; cancer; tumour; vaccine; immunogen; clone.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc.difference 118 /note="undefined residue"
PN WO9837093-A2.
PD 27-AUG-1998.
PR 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (COR1-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PS used in a vaccine for the treatment of prostate cancer
PS Example 3, Page 109; 130pp; English.
CC The present sequence is an immunogenic portion of a prostate tumour
CC protein. The immunogen, or the DNA encoding it, can be used as a
CC vaccine for the treatment of prostate cancer. The immunogen was
CC isolated from a prostate tumour cDNA library obtained by subtracting
CC a prostate tumour cDNA expression library with a normal tissue cDNA
CC library.
SQ Sequence 164 AA:

```

```

alignment_scores:
Quality: 566.00 Length: 139
Ratio: 4.678 Gaps: 2
Percent Similarity: 87.050 Percent Identity: 82.014
alignment_block:
US-09-030-606-175 x W71873 ..

```

```

Align seg 1/1 to: W71873 from: 1 to: 164
28 ATGCAAAACGAATGTCTGCTCGGGCGTCTGTCGATCCGACAGTGGGT 77
|||||
1 MetLunsnGluLeuPheCysSerGlyValLeuValHisProGlnTrpVa 17
78 GCTGTGACCGCAGACAGTGTTCAGAGAACTCTACACCATCGGGCTGGCC 127
|||||
17 IbeuSerAlaAlaHisCysPheGlnAsnSerTyrThrIleGlyLeuGlyL 34

```

```

128 TGCAAGTCTTGAGCGCAGACAGGACGACGAGATGGTGGAGGCC 177
|||||
34 euHISserLeuGluAlaAspGlnGluProGlySerIleMetValGluAla 50
178 AGCCTCCGCTACGAGCACCAGATCAACAGACATCTTGCTGCTACAGA 227
|||||
51 SerLeuSerValAlaArgHisProGluTyrAsnArgProLeuLeuAlaAsnAs 67
228 CCGATCATCTATCAAGTTGGACGAATCCGTCGAGTCTGACACCATCC 277
|||||
67 PluMetLeuIleLysLeuAspGluSerValSerGluSerAspThrIleA 84
278 GGAGCATCAGCATTCGTCGACAGTCCCTACCCGGGGAACTCTTGCCCTC 327
|||||
84 rgserrIleSerIleAlaSerGlnCysProThrAlaGlyAsnSerCysLeu 100
328 GNTCTGGCTGGGGCTGCTGCGGCAACGGCAAGTCCCTACCGTCTGCA 377
|||||
101 ValSerGlyTrpGlyLeuLeuAlaAsnAspAlaVal..... 112
378 CTGCGTGAACGTGCGGTGCTGTGAGGANGTC.....TGCA 415
113 .....IleAlaIleGlnSer***ThrValGlyGlyTrpGluCysG 126
416 GTAAGCTCTATGACCCG 432
126 IuylsLeuSerGlnPro 131
seq_name: A_Geneseq_36:W94493

```

```

seq_documentation_block:
ID W94493 standard; Protein: 268 AA.
AC W94493:
DT 23-APR-1999 (first entry)
DE Human kallikrein.
KW Human; kallikrein; keratinocyte; HKaLL; skin disorder; cancer; eczema;
KW psoriasis; scleroderma; adenocarcinoma; leukaemia; melanoma.
OS Homo sapiens.
PN WO9842849-A1.
PD 01-OCT-1998.
PR 25-MAR-1998; U05939.
PR 26-MAR-1997; US-824874.
PA (INCY-) INCYTE PHARM INC.
PI Hillman JL, Lal P;
DR WPI: 99-070073/06.
DR N-PSDB: X16295.
PT Human kallikrein polypeptide, HKaLL - useful e.g. to treat skin
PT disorders e.g. eczema, psoriasis and screen for antagonists useful
PT to treat skin disorders and cancers
PS Claim 1; Page 41-42; 61pp; English.
CC The present sequence represents human kallikrein, designated HKaLL.
CC HKaLL (or fragments) can be administered therapeutically to increase
CC proteolysis and subsequent skin scaling to treat/prevent skin disorders,
CC e.g. eczema, psoriasis and scleroderma. HKaLL has chemical and
CC structural homology with human stratum corneum chymotryptic enzyme
CC (thought to be involved in the proteolysis of intercellular cohesive
CC structures necessary for desquamation, the process by which outer layers
CC of skin are eliminated), and its expression in cDNA libraries was
CC associated with tumour-associated tissues and skin cells. It can be used
CC to screen for antagonists and agonists, and to generate antibodies.
CC HKaLL antagonists can be used to suppress excessive proteolysis and
CC subsequent skin cell scaling, so may be administered to treat skin
CC disorders. They may also be used to suppress excessive cell
CC proliferation, so can be administered to treat/prevent cancer, e.g.
CC adenocarcinoma, leukaemia and melanoma. Antibodies specific for HKaLL
CC may be used directly as antagonists, or indirectly as a targeting or
CC delivery mechanism for bringing pharmaceutical agents to HKaLL-expressing
CC cells. They are also useful to diagnose conditions/diseases characterised
CC by HKaLL expression and to monitor therapeutic interventions. The
CC polynucleotide encoding HKaLL, or complementary sequences, can be used to
CC produce hybridisation probes, useful to detect polynucleotides encoding
CC HKaLL, e.g. to diagnose diseases relating to polypeptide expression

```

CC (e.g. cancers of the bladder, prostate) or monitor HKAL1 regulation
 CC during therapeutic intervention. Polynucleotides encoding HKAL1 are
 CC useful to produce antisense sequences for therapeutic administration to
 CC modulate/prevent HKAL1 expression e.g. to treat/prevent skin disorders
 CC or cancer as above.
 SQ Sequence 268 AA:

alignment_scores: Length: 213
 Quality: 553.00
 Ratio: 3.393
 Gaps: 3
 Percent Similarity: 76.526 Percent Identity: 48.826

alignment_block:
 US-09-030-606-175 x W94493

Align seg 1/1 to: W94493 from: 1 to: 268

```

13 CAGCGGCGACTGTCATGAA...AACGAATTGTTCTGCGGGCTCT 59
|||||.....:|||||.....:|||||.....:|||||
56 GlnAlaAlaLeuLeuArgProAsnGlnLeuYrCysGlyAlaValLe 72
60 GGTGATCCGAGTGGTGTCTGTCAGCGGACACTGTTCCAGAACTCT 109
|||||.....:|||||.....:|||||.....:|||||
72 uValHisProGlnTrpLeuLeuThrAlaAlaHisCysArgIysValP 89
110 AACCATGGGCTGGCGCTGCAAGTCTTGAGCGGACCAAGAGCCAGG 159
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
89 heArgValArgLeuGlnIstYrSerLeuSerProValYrGlnSerGly 105
160 AGCAGATGTGTGAGGCGCCCTCTCCGTACGCGACCCAGATACACAG 209
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
106 GlnGlnMetPheGlnGlyValIysSerLeuProHisProGlyTyrSerH 122
210 ACTCTGCTGCTAACGACCTCATGCTCATCAAGTGGACGAACTCTGT 259
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
122 sProGlyHisSerAsnAspLeuMetLeuIleLysLeuAsnArgAlaIle 139
260 CCGAGTGTGACACCATCCGAGACATGAGCATGTTGCGAGGCCCTTAC 309
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
139 rGrProThrIysAspValArgProIleAsnValSerSerHisCysProSer 155
310 GCGGGGAACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 359
|||||.....:|||||.....:|||||.....:|||||
156 AlaGlyThrIysCysLeuValSerGlyTrpGlyThrIysSerProGly 172
360 AATG.....CCTACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 403
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
172 nValHisPheProLysValLeuGlnCysLeuAsnIleSerValLeuSerG 189
404 AGGANGTCTGAGTAAAGCTTATGACCCGCTGTACACCCAGCCAGCAT 453
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
189 InLysArgCysGlnAspAlaTrpProAlaGlnIleAspAspThrMetPhe 205
454 TGCCTGCGGAGGAGGAGACAGACAGAGACTCTGCAAGGCTGACTGTG 503
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
206 CysAla...GlyAspIysAlaGlyArgAspSerCysGlnGlyAspSerG 221
504 GGGGCGCTGATCTGCAAGGGGTACTGTCAGAGGCGCTTGTCTTGGAA 553
|||||.....:|||||.....:|||||.....:|||||
221 yGlyProValAlaCysAsnGlySerLeuGlnGlyLeuValSerTrpGly 238
554 AAGCGCGTGGGCAACTGGCGGAGGCTGATGCTTACACCAACTCTGCG 603
|||||.....:|||||.....:|||||.....:|||||
238 sPyrTrpCysAlaArgProAsnArgProGlyValYrThrAsnLeuCys 254
604 AAATCTACTGAGTATAGAGAAAACCTCCAGNCCACT 642
|||||.....:|||||.....:|||||.....:|||||
255 LysPheThrIysTrpIleGlnGlnThrIleGlnAlaAsn 267
seq_name: A_Geneseq_36:R67888
seq_documentation_block:

```

ID R67888 standard; Protein: 253 AA.
 AC R67888;
 DT 09-AUG-1995 (first entry)
 DE Human stratum corneum chymotryptic recombinant enzyme (SCCE)
 KW Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;
 KM callusities; Keratosis pilaris; Ichthyoses; eczema.
 OS Homo sapiens.
 PN MO9500651-A.
 PD 05-JAN-1995.
 PE 20-JUN-1994; IB0166.
 PR 18-JUN-1993; DK-000725.
 PA (SYMB-) SYMBICOM AB.
 PI Egelrud T, Hansson L;
 DR WPI: 95-052088/07.
 DR N-PSDB: 081203.
 PT Nucleotide sequences encoding stratum corneum chymotryptic enzyme
 PT - and related vectors, transformed cells and polypeptides,
 PT useful for treating skin disorders, e.g. acne or psoriasis, and
 PS for identification of specific inhibitors.
 PS Disclosure: Page 97; 137pp; English.
 CC The enzyme encoded by this sequence is used in pharmaceutical, cosmetic
 CC and skin care products, especially to treat and prevent acne,
 CC xeroderma, or other hyperkeratotic conditions (e.g. callusities or
 CC keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is
 CC produced recombinantly following mammal, insect, plant, or
 CC microorganism transformation with plasmid pS507.
 SQ Sequence 253 AA:

alignment_scores: Length: 214
 Quality: 530.50
 Ratio: 3.255
 Gaps: 3
 Percent Similarity: 76.168 Percent Identity: 47.196

alignment_block:
 US-09-030-606-175 x R67888

Align seg 1/1 to: R67888 from: 1 to: 253

```

1 GCGGAGCCCTGGGCGGCGGACATGATGAAACGAATTGTTCTCTC 50
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
40 SerHisProTrpGlnValAlaLeuLeuSerGlyAsnGlnLeuHisCysG 56
51 GGGGCTCTGTTGATCCGCGAGTGGTGTCTGTCAGCGGACACTGTTCC 100
|||||.....:|||||.....:|||||.....:|||||
56 yGlyValLeuValAsnGlnLysTrpValLeuThrAlaAlaHisCysLys 73
101 AGAATCTCTACACCATCGGCTGGGCTGCGACAGTCTTGGAGCGGACAA 150
|||||.....:|||||.....:|||||.....:|||||
73 eHisGlnLysThrValHisLeuGlySerAspThrLeu...GlyAspArg 88
151 GAGCCAGGAGCCAGATGTGTGAGGCGGCGCTCTCCGTACGCGACCCAGA 200
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
89 Arg.....AlaGlnArgIleLysAlaSerLysSerPheArgHisProGly 103
201 GTACACAGACTCTGCTGCTGCTACGACCTCATGCTCATCAAGTGGAGC 250
|||||.....:|||||.....:|||||.....:|||||
103 YrTrpSerThrGlnThrHisValAsnAspLeuMetLeuValLysLeuAsn 120
251 AATCGTGTCCGAGTCTGACACCATCCGAGACATGACATGCTGCTGCGAG 300
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
120 eArgAlaAlaArgLeuSerSerMetValLysLysValArgLeuProSerArg 136
301 TGCCTTACCGGGGGAACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 350
|||||.....:|||||.....:|||||.....:|||||
137 CysGlnLysProGlyThrTrpCysThrValSerGlyTrpGlyThrThr 153
351 GAAGGCG.....AGAATGCTTACCGGCTGCTGCTGCTGCTGCTGCTG 394
|||||.....:|||||.....:|||||.....:|||||
153 rSerProAspValThrPheProSerAspLeuMetCysValAspValLysL 170
395 TGTGTCTGAGGANGTCTGCGATGAGCTTATGACCCGCTGTGTACACCCC 444
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


alignment_scores:

Quality: 498.50 Length: 216
Ratio: 2.950 Gaps: 4
Percent Similarity: 78.241 Percent Identity: 41.667

alignment_block:

US-09-030-606-175 x W10694 ..

Align seg 1/1 to: W10694 from: 1 to: 260

```

1 GCGGAGCCCTGGACGGCCGACCTGTCATGGAAGAAATGTTCTGCTC 50
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
43 SerGlnProTrpGlnAlaAlaLeuPheGlnGlyAlaGlyLeuIleCysG1 59
51 GGGGCTGCTGGTGCATCCGACGGGCTGTCAGCCGACGACCTTTCC 100
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
59 YGlyValLeuValGlyAspArgTrpValLeuThrAlaAlaHisCysLysL 76
101 AGAACTCTTACACATCGCGGCTGACACAGTCTTGAGCCGACCAA 150
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
76 YSGlnLysTyrSerValAlaTrpLeuGlyAspHisSerLeuGlnSerAla 92
151 GAGCCAGGAGCCAGATGTGTGAGCCGACCTCTCGTACGACCCACA 200
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
93 GlnPro...GlnGlnGlnIleGlnValAlaGlnSerIleGlnHisProCy 108
201 GTACACAGACTCTTG.....CTGCTACGACCTCATGCTATCA 241
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
108 sTyraAsnSerAsnProGlnAspHisSerHisAspIleLeuIleA 125
242 AGTTGAGCAATCCGTGTCGAGTGCACACCATCGGAGCATGACATT 291
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
125 rGleuGlnAsnSerAlaAsnLeuGlyAspLysValLysProValGlnLeu 141
292 GCTTCGCAAGTCCCTACCGCGGAGCACTTGCCTGCTGCTGCTGGG 341
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
142 AlaAsnLeuCysProLysValGlyGlnLysCysIleIleSerGlyTrpG1 158
342 TCTGCTGGCAAGCGCAGA.....ATGCTACGCTGCTGACCTGCGGA 385
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
158 yThrValThrSerProGlnGlnAsnProAsnThrLeuAsnGlyAlaG 175
386 ACGTGTGCGTGGTGTCTGAGGAGTCTGCAGTACCTCTATGACCCGCTG 435
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
175 luValLysIleLysSerGlnAsnLysCysGluAlaGlyAlaTyrProGly 191
436 TACCAACCCAGATGTTCTGCGCGCGGAGGCGACAGACAGAGACTC 485
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
192 IleThrGlnGlyMetValCysAla...GlySerSerAsnGlyAlaAsp 207
486 CTGCAACGGTGAATCTGGGGGGCCCTGATCTGCAACGGGTACTGCA 535
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
207 rCysGlnIleLysPheSerGlyGlyProLeuValCysAspIleMetLeu 224
536 GCCTGTGTCTTTGGAAAGCCCGGTGTGGCCACTTGCGCTGCCAGGT 585
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
224 YLleThrSerTrpLysSerAspProCysGlyLysProGlnLysProGly 240
586 GTCTACACCAACTCTGCAATTCCTACTAGTGGATAGAGAAACCGCT 633
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
241 ValTyrThrLysIleCysArgTyrThrThrTrpIleLysLysThrMet 256

```

seq_name: A_Geneseq_36:W12393

seq_documentation_block:

ID W12393 standard; Protein; 260 AA.
AC W12393;
DT 15-MAY-1997 (first entry)
DE Mouse neuropsin protein.
KW Mouse; neuropsin; hippocampus; lambda gt10; primer; PCR; amplification;
polymerase chain reaction; serine protease domain; nerve growth factor;
NGF; insect cell; virus; expression vector; transfection;
KW cerebral disease.

OS Mus musculus.
PN J08311099-A.
PD 26-NOV-1996.
PF 13-MAR-1996: 056367.
PR 14-MAR-1995: JP-054584.
PA (SHIO/) SHIOZAKA S.
DR WPI: 97-061812/06.
DR N-PSDB: T63251.
PT Nucleic acid encoding neuropsin - for producing neuropsin, useful
for diagnosis and treatment of cerebral disease
PS Claim 1: Page 6-7, 9pp; Japanese.
CC This is the amino acid sequence of a novel mouse protein designated
CC neuropsin. The encoding gene was isolated from a mouse hippocampal
CC cDNA library in lambda gt10 using a cloned, amplified fragment of
CC the gene (clone B41: T63254). This fragment was amplified using
CC primers T63252-3. The primers were synthesised based on the serine
CC protease domain of nerve growth factor (NGF)-gamma. The screen isolated
CC 6 positive clones, of which clone NP5 contained the longest insert
CC (this sequence). The protein has a molecular weight of around 26 kD.
CC It has 43% homology with EGF-BP, 41 % with NGF-gamma; 39% with NGF-alpha;
CC 38% with trypsin and 18% with tPA. The protein can be used for clinical
CC diagnosis and treatment of cerebral diseases.
SQ Sequence 260 AA;

alignment_scores:

Quality: 498.50 Length: 216
Ratio: 2.950 Gaps: 4
Percent Similarity: 78.241 Percent Identity: 41.667

alignment_block:

US-09-030-606-175 x W12393 ..

Align seg 1/1 to: W12393 from: 1 to: 260

```

1 GCGGAGCCCTGGACGGCCGACCTGTCATGGAAGAAATGTTCTGCTC 50
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
43 SerGlnProTrpGlnAlaAlaLeuPheGlnGlyAlaGlyLeuIleCysG1 59
51 GGGGCTGCTGGTGCATCCGACGGGCTGTCAGCCGACGACCTTTCC 100
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
59 YGlyValLeuValGlyAspArgTrpValLeuThrAlaAlaHisCysLysL 76
101 AGAACTCTTACACATCGCGGCTGACACAGTCTTGAGCCGACCAA 150
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
93 GlnPro...GlnGlnGlnIleGlnValAlaGlnSerIleGlnHisProCy 108
201 GTACACAGACTCTTG.....CTGCTACGACCTCATGCTATCA 241
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
108 sTyraAsnSerAsnProGlnAspHisSerHisAspIleLeuIleA 125
242 AGTTGAGCAATCCGTGTCGAGTGCACACCATCGGAGCATGACATT 291
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
125 rGleuGlnAsnSerAlaAsnLeuGlyAspLysValLysProValGlnLeu 141
292 GCTTCGCAAGTCCCTACCGCGGAGCACTTGCCTGCTGCTGCTGGG 341
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
142 AlaAsnLeuCysProLysValGlyGlnLysCysIleIleSerGlyTrpG1 158
342 TCTGCTGGCAAGCGCAGA.....ATGCTACGCTGCTGACCTGCTGA 385
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
158 yThrValThrSerProGlnGlnAsnProAsnThrLeuAsnGlyAlaG 175
386 ACGTGTGCGTGGTGTCTGAGGAGTCTGCAGTACCTCTATGACCCGCTG 435
  :::::::::::::::::::::::::::: : : : : : : : : : : : :
175 luValLysIleLysSerGlnAsnLysCysGluAlaGlyAlaTyrProGly 191
436 TACCAACCCAGATGTTCTGCGCGCGGAGGCGACAGACAGAGACTC 485
  :::::::::::::::::::::::::::: : : : : : : : : : : : :

```

```

192 IleThrGluIleMetValCysAla...GlySerSerAsnGlyAlaSerThr 207
486 CTCACAGCGACTCTGGGGGCCCCGATGCTGCAAGCGGCTACTGTGACG 535
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 rCysGlnGlyAspSerGlyIleProLeuValCysAspGlyMetLeuGln 224
536 GCGTGTGTCTTTCGGAAGACCCGCTGGCCAACTTGGCGTCCGAGT 585
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
224 lYleThrSerTrpGlySerAspProCysGlyLysProGluLysProGly 240
586 GGTACACCACTCTGCAATTCAGTGAATGATAGAGAAACCGTC 633
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 ValTrpThrLysIleCysArgTrpThrTrpIleLysLysThrMet 256
seq_name: A_Geneseq_36:R44532

```

```

seq_documentation_block:
ID R44532 standard; Protein; 244 AA.
AC R44532;
DR 01-JUL-1994 (first entry)
DE Zyme APP-cleaving protease.
KM Amyloid precursor protein-cleaving protease; Alzheimer's disease;
  Down's syndrome; diagnosis; propensity.
OS Homo sapiens.
PN EP-576152-A.
PD 29-DEC-1993.
PE 26-MAY-1993; 304103.
PR 28-MAY-1992; US-891542.
PA (Ertl & CO Ltd.
PI Dixon BP, Johnstone EM, Little SP, Norris FH;
  WPI: 94-001306/01.
DR N-PSDB: Q53487.
PT Amyloid precursor protein-cleaving protease - associated with
  Alzheimer's disease and Down's syndrome
PS Claim 1: Page 11-12: 14pp; English.
CC The sequence is that of Zyme, an amyloid precursor protein-cleaving
  protease. It cleaves APP to generate amyloidogenic fragments of the
  size expected of a Met596-Asp597 cleavage. It is thus very useful
  in furthering the characterisation of Alzheimer's disease and
  Down's syndrome.
CC Down's syndrome.
SQ Sequence 244 AA:

```

```

alignment_scores:
Quality: 470.50 Length: 215
Ratio: 3.055 Gaps: 3
Percent Similarity: 71.628 Percent Identity: 42.326.

```

alignment_block:

US-09-030-606-175 x R44532 ..

Align seg 1/1 to: R44532 from: 1 to: 244

```

1 GCCGACGCCCTGGCAGGGGCGACTGTGATGAAACGAATTTGTTCTC 50
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
32 SerHisProTrpGlnAlaIleLeuTrpHisSerGlyHisLeuLeuGly 48
51 GGGCGTCTGCTGTCATCCGAGTGGTGTCTGAGCCGACACATGTTTCC 100
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48 yGlyValLeuIleHisProLeuTrpValLeuThrAlaHisCysLysL 65
101 AGAAGTCTTACACCATGGGCTGGCGCTGCACATCTT.....GAG 141
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
65 ySPProAsnLeuGlnValPheLeuGlyLysHisAsnLeuArgGln 81
142 GCCGACCAAGACCCAGGAGCCAGATGTGAGCCGACCTTCCTCGTAC 191
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
82 SerSerGlnGlnGln...SerSerValValArgAla.....ValI 94
192 GCACCCAGAGTACACAGACTCTTGTCTGCTAAGACCTCATGCTCATCA 241
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
94 eHisProAspTrpAspAlaIleSerHisAspGlnAspIleMetLeu 111

```

```

242 AGTTGACCAATCCGTGTCGATGTCGACACCATCCGAGCATGACAT 291
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
111 rGluLeuAlaTrpProAlaLysLysSerGlyLeuIleGlnProLeuProLeu 127
292 GCTTCCCAATGCCCTACCGCGGGAACCTGTGCTGCTGCTGCTGCTG 341
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
128 GluArgAspCysSerLysAlaSerHisThrSerCysHisIleLeuGly 144
342 TCTGCTGAGGACCGGACCAATGCTACGCTGCTGCTGCTGCTGCTG 391
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
144 yLysThrAlaAspGlyAspPheProAspThrIleGlnCysAlaTrp 161
392 CGGTGCTGTGTCGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 441
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
161 lSerValSerArgGluGluCysGlnHisAlaTrpProGlyIleThr 177
442 CCCAGCATGTTTGCSCGCGGAGGCGACAGCAGACAGACTCTGCAA 491
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
178 GluAsnMetLeuCysAlaGlyAspGlyLysTrpGlyLysAspSerCys 194
492 CGGTGCTGTGTCGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 541
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
194 nGlyAspSerGlyGlyProLeuValCysGlyAspHisLeuArgGlyLeu 211
542 TGTCTTTCGGAAGACCCGCTGGCCAACTTGGCGTCCGAGTGTCTAC 591
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
211 alSerTrpGlyAsnIleProCysGlySerLysGlyLysProGlyValTrp 227
592 ACCAAGCTCTGCAATTCAGTGAATGATAGAGAAACCGTCCAG 636
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
228 ThrAsnValCysArgTrpThrAsnTrpIleGlnLysThrIleGln 242
seq_name: A_Geneseq_36:W22985

```

seq_documentation_block:

ID W22985 standard; Protein; 244 AA.

AC W22985;

DR 08-OCT-1997 (first entry)

DE Human serine protease 59 (SP59).

KM Human: colon carcinoma; COLO 201; cell line; serine protease; SP59;

KW screening; inhibitor; treatment; disease.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Peptide 1..21

FT Peptide /label= sig_peptide

FT Peptide 22..244

FT Peptide /label= mat_peptide

FN J09149790-A.

PD 10-JUN-1997.

PE 24-JUL-1996; 212196.

PR 29-SEP-1995; JP-275105.

PA (SUNR) SUNTORY LTD.

DR WPI: 97-357902/33.

DR N-PSDB: T79126.

PT Human colon carcinoma derived serine protease(s) SP59, SP60 and SP67

PT - useful to screen for specific inhibitors, e.g. to search for, or

PT study agent for treatment of various diseases

PS Claim 1: Pages 10-11; 16pp; Japanese.

CC The present sequence is the human colon carcinoma COLO 201

CC cell line derived serine protease 59 (SP59), which can be used to

CC screen for specific inhibitors, e.g. to search for, or study an

CC agent for the treatment of various diseases.

SQ Sequence 244 AA:

```

alignment_scores:
Quality: 470.50 Length: 215
Ratio: 3.055 Gaps: 3
Percent Similarity: 71.628 Percent Identity: 42.326

```

alignment_block:

US-09-030-606-175 x W22985 ..

Align seg 1/1 to: W22985 from: 1 to: 244

```
1 GCGCAGCCCTGGCAGCGCGCACTGGTCATGGAAGAAATGTTCTGCTC 50
   ::::::::::::::::::::
32 SerHisProTyrGlnAlaLeuTyrThrSerGlyHisLeuLeuCysG1 48
   ::::::::::::::::::::
51 GGGCGTCTGTGTCATCCGCGCACTGGTGTGTCAGCCGCACTGTTC 100
   ::::::::::::::::::::
48 yGlyValLeuIleHisProLeuTyrValLeuThrAlaHisCysLysL 65
   ::::::::::::::::::::
101 AGAATCTCTACACATCGGGCTGGCGCTGCACATCTT.....GAG 141
   ::::::::::::::::::::
65 ySPProAsnLeuGlnValPheLeuGlyLysHisAsnLeuArgGlnArg 81
   ::::::::::::::::::::
142 GCCGACCAAGAGCCAGGAGCGAGATGGTGGAGCCAGCCCTCCGTCAG 191
   ::::::::::::::::::::
82 SerSerGlnGln.....SerSerValAlaArgAla.....ValI1 94
   ::::::::::::::::::::
192 GCACCCAGATACACAGACATCTTGTCTGCTACAGACCTCATGTATCA 241
   ::::::::::::::::::::
94 eHisProAspTyrAlaAlaSerHisAspGlnAspIleMetLeuAla 111
   ::::::::::::::::::::
242 AGTGGAGAGATCCGTGTCGCGAGTGTGACACATCCGAGCATCAGCAT 291
   ::::::::::::::::::::
111 rGleuAlaArgProAlaLysLeuSerGluLeuIleGlnProLeuProLeu 127
   ::::::::::::::::::::
292 GCTTCGACGATCCCTACCGCGGGAATCTTGCCTGCTGCTGGGAG 341
   ::::::::::::::::::::
128 GluArgAspCysSerAlaAsnThrThrSerCysHisIleLeuGlyTyrP 144
   ::::::::::::::::::::
342 TCTGCTGGCGAAGCGCAGATGCTTACCGTGTGCTGCTGCAACGTGT 391
   ::::::::::::::::::::
144 yLysThrAlaAspGlyAspPheProAspThrIleGlnCysAlaTyrIle 161
   ::::::::::::::::::::
392 CGGTGTGTCTGAGGANGTCTGCATAGCTATAGCCCGCTGTACCAAC 441
   ::::::::::::::::::::
161 IsLeuValSerArgGlnGlnCysGluHisAlaTyrProGlyGlnIleThr 177
   ::::::::::::::::::::
442 CCCAGCATGTTCTGGCGCGCGAGGCAAGCAAGCAAGGACTCTGCA 491
   ::::::::::::::::::::
178 GlnAsnMetLeuCysAlaGlyAspGlnLysTyrGlyLysAspSerCysG 194
   ::::::::::::::::::::
492 CGGTGACTCTGGGGGCGCCCTGATGTGCAACGGGTACTTGCAAGGCC 541
   ::::::::::::::::::::
194 nGlyAspSerGlyGlyProLeuValCysGlyAspHisLeuArgGlyLeu 211
   ::::::::::::::::::::
542 TGTCTTTGGAAAGCCCGGTGTGCGCAACTTGGCGTGCAGAGTGTAC 591
   ::::::::::::::::::::
211 aLysTyrPglYAsnIleProCysGlySerLysGluLysProGlyValTyr 227
   ::::::::::::::::::::
592 ACCAACCCTGTGCAATTCATAGTGAATAGAGAAACCGTCCAG 636
   ::::::::::::::::::::
228 ThrAsnValCysArgTyrThrAsnIleGlnLysThrIleGln 242
```

This Page Blank (uspto)


```

DB 361 ATGCTTACCGTCTGACACGCGGAAACGTCGTCGTGCTGAGGAGNCTGCACTAAG 420
QY 421 CTTATGACCCCGGTGACACCCGACATGTTCTGCGCGGGGAGGGAAGCCAGAG 480
DB 421 CTTATGACCCCGGTGACACCCGACATGTTCTGCGCGGGGAGGGAAGCCAGAG 480
QY 481 GACTCTGCAAGGTCAGTCTGCGGGGGCCCTGATTCGACAGGGTACTTGCAGGGCTT 540
DB 481 GACTCTGCAAGGTCAGTCTGCGGGGGCCCTGATTCGACAGGGTACTTGCAGGGCTT 540
QY 541 GTGTCTTTGGGAAAGCCCGGTGTGCGCAACTTGGCGTCCAGGTCTTACACCACTTC 600
DB 541 GTGTCTTTGGGAAAGCCCGGTGTGCGCAACTTGGCGTCCAGGTCTTACACCACTTC 600
QY 601 TGCAAATTCAGTGTAGTAGAAGAACGTCACAGTCACTTACCTGCGGGAGTGGGA 660
DB 601 TGCAAATTCAGTGTAGTAGAAGAACGTCACAGTCACTTACCTGCGGGAGTGGGA 660
QY 661 ACCCATGAATTTGACCCCAATATCATCTCGGGAANGAATTCAGGAATATCTGTCCCA 720
DB 661 ACCCATGAATTTGACCCCAATATCATCTCGGGAANGAATTCAGGAATATCTGTCCCA 720
QY 721 GCCCTCTCTCTCTCAGGCGGAGAGTCCAGGCGCCCTCTCTCTCTCTCTCTCTCTCT 780
DB 721 GCCCTCTCTCTCTCAGGCGGAGAGTCCAGGCGCCCTCTCTCTCTCTCTCTCTCTCT 780
QY 781 GTACAGATCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840
DB 781 GTACAGATCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840
QY 841 CCNTCAGACCCAGAGATCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
DB 841 CCNTCAGACCCAGAGATCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
QY 901 CCNTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
DB 901 CCNTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
QY 961 TCCAAAGCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
DB 961 TCCAAAGCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
QY 1021 TCAGACCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
DB 1021 TCAGACCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
QY 1081 NGTTGACCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1140
DB 1081 NGTTGACCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1140
QY 1141 ATTAAGTNTAAGAGAGCCCAAAAAA 1167
DB 1141 ATTAAGTNTAAGAGAGCCCAAAAAA 1167

```

RESULT 2

```

ID V61252 standard; cDNA: 1167 BP.
AC V61252:
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE13.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.
PD 27-AUG-1998.
PE 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J.
DR WPI; 98-609886/51.

```

```

DR P-PSDB: W71782.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 107; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;

Query Match          98.6%; Score 1151; DB 1; Length 1167;
Best Local Similarity 100.0%; Pred. No. 2,7e-278;
Matches 1167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCAGCCCTGCGAGCGGCGACTGTGTAAGAAAGAAATGTTGCTGCGGCGTCTG 60
DB 1 GCGCAGCCCTGCGAGCGGCGACTGTGTAAGAAAGAAATGTTGCTGCGGCGTCTG 60
QY 61 GTGCATCCGCACTGGGTGCTGTGACCCGACACTGTTCGAAACTCTTACACCACTG 120
DB 61 GTGCATCCGCACTGGGTGCTGTGACCCGACACTGTTCGAAACTCTTACACCACTG 120
QY 121 CTGGGCTCTGACAGTCTTGAGGCGCAGCAAGAGCCAGGAGCCAGATGTGAGGCGCAGC 180
DB 121 CTGGGCTCTGACAGTCTTGAGGCGCAGCAAGAGCCAGGAGCCAGATGTGAGGCGCAGC 180
QY 181 CTCTCCGTAAGCCACCCAGATGACAGACTCTTGTCTGCTACGACCTCATGCTATC 240
DB 181 CTCTCCGTAAGCCACCCAGATGACAGACTCTTGTCTGCTTAAAGACTCATGCTATC 240
QY 241 AAGTTGGAAGATTCGTCGTCGAGTGTGACACATCCGAGATGAGATGCTTCGAG 300
DB 241 AAGTTGGAAGATTCGTCGTCGAGTGTGACACATCCGAGATGAGATGCTTCGAG 300
QY 301 TCCCTTACCGCGGGGAACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
DB 301 TCCCTTACCGCGGGGAACCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
QY 361 ATGCTTACCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
DB 361 ATGCTTACCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
QY 421 CTTATGACCCCGGTGACACCCGACATGTTCTGCGCGGGGAGGGAAGCCAGAG 480
DB 421 CTTATGACCCCGGTGACACCCGACATGTTCTGCGCGGGGAGGGAAGCCAGAG 480
QY 481 GACTCTGCAAGGTCAGTCTGCGGGGGCCCTGATTCGACAGGGTACTTGCAGGGCTT 540
DB 481 GACTCTGCAAGGTCAGTCTGCGGGGGCCCTGATTCGACAGGGTACTTGCAGGGCTT 540
QY 541 GTGTCTTTGGGAAAGCCCGGTGTGCGCAACTTGGCGTCCAGGTCTTACACCACTTC 600
DB 541 GTGTCTTTGGGAAAGCCCGGTGTGCGCAACTTGGCGTCCAGGTCTTACACCACTTC 600
QY 601 TGCAAATTCAGTGTAGTAGAAGAACGTCACAGTCACTTACCTGCGGGAGTGGGA 660
DB 601 TGCAAATTCAGTGTAGTAGAAGAACGTCACAGTCACTTACCTGCGGGAGTGGGA 660
QY 661 ACCCATGAATTTGACCCCAATATCATCTCGGGAANGAATTCAGGAATATCTGTCCCA 720
DB 661 ACCCATGAATTTGACCCCAATATCATCTCGGGAANGAATTCAGGAATATCTGTCCCA 720
QY 721 GCCCTCTCTCTCTCAGGCGGAGAGTCCAGGCGCCCTCTCTCTCTCTCTCTCTCTCT 780
DB 721 GCCCTCTCTCTCTCAGGCGGAGAGTCCAGGCGCCCTCTCTCTCTCTCTCTCTCTCT 780
QY 781 GTACAGATCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840
DB 781 GTACAGATCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840

```

QY	841	CCNCTACAGCCAGAGAGTCCAGCCCTCTCCTCCTCAGCAGAGAGTCCAGAGCCCCCAGC	900
Db	841	CCNCTACAGCCAGAGAGTCCAGCCCTCTCCTCCTCAGCAGAGAGTCCAGAGCCCCCAGC	900
QY	901	CCNCTCCTCCTCAGAGCCAGAGGGTGCAGAGCCCCCAACCCCTCTCCTCAGAGTCCAGAG	960
Db	901	CCNCTCCTCCTCAGAGCCAGAGGGTGCAGAGCCCCCAACCCCTCTCCTCAGAGTCCAGAG	960
QY	961	TCCAGAGCCCCCAACCCCTCTCTCCCGCAGAGCCAGAGTCCAGAGTCCAGAGCCCCCTCTCC	1020
Db	961	TCCAGAGCCCCCAACCCCTCTCTCCCGCAGAGCCAGAGTCCAGAGTCCAGAGCCCCCTCTCC	1020
QY	1021	TCAAGCCAGCCAGCGGTCCCAATCCACCTAGANTNTCCCTGTACACAGATGCCCTCTGTGCA	1080
Db	1021	TCAAGCCAGCCAGCGGTCCCAATCCACCTAGANTNTCCCTGTACACAGATGCCCTCTGTGCA	1080
QY	1081	NGTTCAGCCCAACCTTACCAATTTGGTTTTCATTTTTTGTCCCTTCCCTTCAGATTCAGAA	1140
Db	1081	NGTTCAGCCCAACCTTACCAATTTGGTTTTCATTTTTTGTCCCTTCCCTTCAGATTCAGAA	1140
QY	1141	ATAAGGTTTAAGAGAAAGCCGAAAAAAA	1167
Db	1141	ATAAGGTTTAAGAGAAAGCCGAAAAAAA	1167

QY	110	ACACCACTGGGGCTGGGGCTCTACAGCTCTTGAGGCGCACCAAGAGCCAGGGACCAAGATGG	169
Db	161	ACACCACTGGGGCTGGGGCTCTACAGCTCTTGAGGCGCACCAAGAGCCAGGGACCAAGATGG	220
QY	170	TGAGAGCCAGCCTCTCCGTACGGGACCCAGAGTACAAACAGACTCTTGTCTCGCTAACGACC	229
Db	221	TGAGAGCCAGCCTCTCCGTACGGGACCCAGAGTACAAACAGACCCTGCTTCGCTAACGACC	280
QY	230	TCATGCTCATCAAGTTGGAGGAATCCGTGTCCGAGTCTGACACCATCCGAGACATCAGCA	289
Db	281	TCAAGCTCATCAAGTTGGAGGAATCCGTGTCCGAGTCTGACACCATCCGAGACATCAGCA	340
QY	290	TTGGTTGGCAGTGGCCCAACCGGGGGGAACTCTGCTCGTTCGCTGGGCTGGGGCTGGCTGG	349
Db	341	TTGGTTGGCAGTGGCCCAACCGGGGGGAACTCTTGGCTTCGCTGGGCTGGGGCTGGCTGG	400
QY	350	CGAAGCGCAGAAATGCTTACCCGTCTGCTGCTGACGTCTGGAAGCTGTGCGTGTCTGAGGANG	409
Db	401	CGAAGCGCAGAAATGCTTACCCGTCTGCTGCTGACGTCTGGAAGCTGTGCGTGTCTGAGGANG	460
QY	410	TTCTGCAATAGCTCTATGACCCGCGCTGTACACCCACAGATGTTCTGCGCCGCGGGAGGGCC	469

RESULT	3
V58644	
ID	V58644 standard; cDNA; 1248 bp.
AC	V58644;
DT	08-DEC-1998 (first entry)
DE	Prostate tumour specific gene clone DEL.
KW	Prostate tumour specific gene; human; prostate cancer; detection; therapy; SS.
OS	Homo sapiens.
FH	Key
FT	Location/Qualifiers
FT	217..696
CDS	/*tag= a
PN	W09837418-A2.
PD	27-AUG-1998.
PF	25-FEB-1998; U03690.
PR	09-FEB-1998; US-904809.
PR	25-FEB-1997; US-806596.
PR	01-AUG-1997; US-904809.
PA	(COR-) CORIXA CORP.
PI	Dillon DC, Xu J;
PI	WPI: 98-480805/41.
DR	P-PSDB: W69387.
PT	Novel human prostate specific tumour protein and fragments - useful
PT	for detecting and treating prostate cancers
PS	Claim 1; Page 112; 141pp; English.
CC	This sequence represents a human prostate tumour specific gene, and can
CC	be used in the method of the invention. The method is for detecting
CC	prostate cancer comprises contacting a biological sample with an agent
CC	able to bind an immunogenic portion of a prostate protein (such as
CC	encoded by this sequence). An antibody which binds to an immunogenic
CC	portion of the prostate protein, and the method can be used to detect,
CC	monitor progression of, or treat prostate cancers. The antibody may
CC	also be conjugated to a therapeutic agent for use in therapy of prostate
CC	cancers.
Sequence	1248 BP; 288 A; 424 C; 303 G; 228 T;

Db	521	AAGGCCAAGAAAGACTCCTGCACGGTAGCTCTGGGGGGCCCTCGATCTCAACGGGTACT	580
QY	530	TGCAGGGCCCTTGATGCTCTTTTCGGAAAAAGCCCGCTGTGGCCAACTTGGCGTCCAGAGTGTCT	589
Db	581	TGCAGGGCCCTTGATGCTCTTTTCGGAAAAAGCCCGCTGTGGCCAAAGTGTGGCGTCCAGAGTGTCT	640
QY	590	ACACCAACCTCTCTGCAAATTTACTGAGTGGATGAGAAAAACCTGCCAGNCCAGTTAACTCT	649
Db	641	ACACCAACCTCTCTGCAAATTTACTGAGTGGATGAGAAAAACCTGCCAGNCCAGTTAACTCT	700
QY	650	GGGAGCTGGGAAACCAATGAATTTAGACCCCAATTAATCTGTGGGAANAATTCAGAGAT	709
Db	701	GGGAGCTGGGAAACCAATGAATTTAGACCCCAATTAATCTGTGGGAANAATTCAGAGAT	760
QY	710	ATCTGTTCCAGGCCCTCTCTCCCTCAGGGCCAGAGATCCAGGGCCCCAGCCCTCTCTCC	769
Db	761	ATCTGTTCCAGGCCCTCTCTCCCTCAGGGCCAGAGATCCAGGGCCCCAGCCCTCTCTCC	820
QY	770	TCAAAACCAAGGATACAGATATCCCAAGCCCTCTCTCTCAGACCCAGAGATCCAGACCC	829
Db	821	TCAAAACCAAGGATACAGATATCCCAAGCCCTCTCTCTCAGACCCAGAGATCCAGACCC	880
QY	830	CAGGCCCTCTCTCTCAGACCCAGAGATCCAGGCCCTCTCTCTCAGAGCAGAGATCCA	889
Db	881	CAGGCCCTCTCTCTCAGACCCAGAGATCCAGGCCCTCTCTCTCAGAGCAGAGATCCA	940
QY	890	GACCCCCAGACCCCTCTCTCTCAGACCCAGAGGATCAGGCCCCCAACCCCTCTCTCTCT	949
Db	941	GACCCCCAGACCCCTCTCTCTCAGACCCAGAGGATCAGGCCCCCAACCCCTCTCTCTCT	1000
QY	950	AGAGTCAGAGATCCAAAGCCCCCAACCCCTCTCTCCCAACCCAGAGATTCAGAGTCCAG	1009
Db	1001	AGAGTCAGAGATCCAAAGCCCCCAACCCCTCTCTCCCAACCCAGAGATTCAGAGTCCAG	1060

[illegible]

Accession	Sequence	Position
Db	CCCCCTCCTCCTAGACCCACGCGGTCCAAATGCACCTAACATCTCCCTGTAACAGTGGC	1120
Db	1061 CCCCCTCCTCCTAGACCCACGCGGTCCAAATGCACCTAACATCTCCCTGTAACAGTGGC	1120
Qy	1070 CCGTGTGGCAGTGTGACCCCAACCTTACCGAGTGTGTTTCATATTTTGTCCCTTCCCC	1122
Db	1121 CCGTGTGGCAGTGTGACCCCAACCTTACCGAGTGTGTTTCATATTTTGTCCCTTCCCC	1180
Qy	1130 TAGATCCAGAAATTAAGTNTTAAGAGAGCGCAAAAAA	1167
Db	1181 TAGATCCAGAAATTAAGTNTTAAGAGAGCGCAAAAAA	1218

RESULT 4

ID V61249 standard; cDNA; 1248 BP.

AC V61249:

DT 06-JAN-1999 (first entry)

DE cDNA sequence of prostate tumour clone P703 splice variant DE1.

KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.

OS Homo sapiens.

PN M09837093-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03492.

PR 09-FEB-1998; US-020956.

PR 25-FEB-1997; US-806099.

PR 01-AUG-1997; US-904804.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J.

DR WPI: 98-60986/51.

P-PSDB: W1871.

PT Polypeptides comprising immunogenic portions of prostate proteins - used in a vaccine for the treatment of prostate cancer

PS Claim 3; Page 104; 130pp; English.

CC The present sequence is a new DNA which encodes an immunogenic portion of a prostate tumour protein. The encoded immunogen, or the DNA itself, can be used as a vaccine for the treatment of prostate cancer. The DNA was identified by analysis of a subtracted cDNA library obtained by CC subtracting a prostate tumour cDNA expression library with a normal CC tissue cDNA library.

Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 95.2%; Score 1111.2; DB 1; Length 1248;

Best Local Similarity 96.7%; Pred. No. 2.3e-268;

Matches 1139; Conservative 0; Mismatches 27; Indels 12; Gaps 1;

QY 2 CGCAGCCCTGGGAGCGGCACTGTCATGGAAGAAATGTTGCTGGCGGCTCTGG 61

DB 41 CGCAGCCCTGGGAGCGGCACTGTCATGGAAGAAATGTTGCTGGCGGCTCTGG 100

QY 62 TGCATCCGAGTGGGAGTGTGTCAGCGGCACTGTTCCAGAA-----CTCCT 109

DB 101 TGCATCCGAGTGGGAGTGTGTCAGCGGCACTGTTCCAGAAAGTGTGCAAGACTCT 160

QY 110 ACACCATCGGCGCTGGGCTGACAGTCTTGAGCGCCAGCAGAGCGGAGCGAGATG 169

DB 161 ACACCATCGGCGCTGGGCTGACAGTCTTGAGCGCCAGCAGAGCGGAGCGAGATG 220

QY 170 TGGAGCGCAGCTCTCCGTCAGCGCACCCAGAGTACAAAGACTCTTGTCTGCTAAAGACC 229

DB 221 TGGAGCGCAGCTCTCCGTCAGCGCACCCAGAGTACAAAGACTCTTGTCTGCTAAAGACC 280

QY 230 TCATGCTCATCAAGTTGGAGCAATCCGTCGTGAGTCAGCAACCAATCCGAGACATCAGCA 289

DB 281 TCATGCTCATCAAGTTGGAGCAATCCGTCGTGAGTCAGCAACCAATCCGAGACATCAGCA 340

QY 290 TTGCTTCGCAAGTCCCTACCGCGGGGAACTCTTGCTCTGTTGCTGGGCTGCTGCTGG 349

DB 341 TTGCTTCGCAAGTCCCTACCGCGGGGAACTCTTGCTCTGTTGCTGGGCTGCTGCTGG 400

QY 350 CGAAGCGCAGAAATGCCCTACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 409

DB 401 CGAAGCGCAGAAATGCCCTACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 460

QY 410 TGTGAGTAGTCTATAGCCGCTGTACACCCAGCATGTTGCGCGCGCGGAGAGGC 469

DB 461 TGTGAGTAGTCTATAGCCGCTGTACACCCAGCATGTTGCGCGCGCGGAGAGGC 520

QY 470 AAGACCAAGAGACTCTGTCAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 529

DB 521 AAGACCAAGAGACTCTGTCAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 580

QY 530 TGCAGGCGCTTGTGCTTTCGAAAAAGCCCGTGTGCGCACTTGCGCTGCTGCTGCTGCT 589

DB 581 TGCAGGCGCTTGTGCTTTCGAAAAAGCCCGTGTGCGCACTTGCGCTGCTGCTGCTGCT 640

QY 590 ACACCAACCTCTGCAAAATTCATGATGATAGAGAAACCGTCACATCCAGTAACTCT 649

DB 641 ACACCAACCTCTGCAAAATTCATGATGATAGAGAAACCGTCACATCCAGTAACTCT 700

QY 650 GGGGACTGGGAACCCATGAATATGACCCCAATATACATCTGCGGAAGAATTCAGGAAT 709

DB 701 GGGGACTGGGAACCCATGAATATGACCCCAATATACATCTGCGGAAGAATTCAGGAAT 760

QY 710 ATCTGTTCCAGCCCT 769

DB 761 ATCTGTTCCAGCCCT 820

QY 770 TCAAACCAAGGATAGATATCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 829

DB 821 TCAAACCAAGGATAGATATCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 880

QY 830 CAGCCCT 889

DB 881 CAGCCCT 940

QY 890 GACCCCT 949

DB 941 GACCCCT 1000

QY 950 AGAGTCAGAGTTCAGAGCCCGCCCAACCCCTGTCCTCCAGACCCAGAGTTCAGTCCAG 1009

DB 1001 AGAGTCAGAGTTCAGAGCCCGCCCAACCCCTGTCCTCCAGACCCAGAGTTCAGTCCAG 1060

QY 1010 CCCCCT 1069

DB 1061 CCCCCT 1120

QY 1070 CCCCCT 1129

DB 1121 CCCCCT 1180

QY 1130 TAGATCCAGAAATTAAGTTAAGTAAGAGCCGCAAAAAA 1167

DB 1181 TAGATCCAGAAATTAAGTTAAGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1218

RESULT 5

V58645

ID V58645 standard; cDNA; 1265 BP.

AC V58645:

DT 08-DEC-1998 (first entry)

DE Prostate tumour specific gene clone DE2.

KW Prostate tumour specific gene; human; prostate cancer; detection; therapy; ss.

OS Homo sapiens.

PN M09837418-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03690.

PR 09-FEB-1998; US-904809.

PR 25-FEB-1997; US-806596.

PR 01-AUG-1997; US-904809.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J.

DR WPI: 98-480805/41.

PT Novel human prostate specific tumour protein and fragments - useful for detecting and treating prostate cancers

PS Claim 1; Page 113-114; 141pp; English.

CC This sequence represents a human prostate tumour specific gene, and can be used in the method of the invention. The method is for detecting prostate cancer comprises contacting a biological sample with an agent able to bind an immunogenic portion of a prostate protein (such as encoded by this sequence). An antibody which binds to an immunogenic portion of the prostate protein, and the method can be used to detect, monitor progression of, or treat prostate cancers. The antibody may also be conjugated to a therapeutic agent for use in therapy of prostate cancers.


```

QY 470 AAGACGAGAGAGACTCCTGCAAGGTGACTCTGGGGGGCCCTGTATCTGCAAGGGTACT 529
    |||
Db 501 AAGACGAGAGAGACTCCTGCAAGGTGACTCTGGGGGGCCCTGTATCTGCAAGGGTACT 560
QY 530 TGCAGGGCCTTGCTGCTTTCGGAAAAAGCCCGCTGTGGCAACTTGGCTGCGAGGTGCT 589
    |||
Db 561 TGCAGGGCCTTGCTGCTTTCGGAAAAAGCCCGCTGTGGCAACTTGGCTGCGAGGTGCT 620
QY 590 ACACCAACCTCTGCAAAATTCAGTGTGATGAGAAAAAGCT-CCAGNCCAGTTAATC 648
    |||
Db 621 ACACCAACCTCTGCAAAATTCAGTGTGATGAGAAAAAGCTGACAGAGAGTAACTC 680
QY 649 TGGGAGCTGGGAGACCCATGAATTTGACCCCAATATCTCTGGGGAAGAATTCAGAA 708
    |||
Db 681 TGGGAGCTGGGAGACCCATGAATTTGACCCCAATATCTCTGGGGAAGAATTCAGAA 740
QY 709 TATCTGTTCCAGAGCCCTCTCTCCCTCAGAGCCAGAGTCCAGGCCCCCTCTCTCC 768
    |||
Db 741 TATCTGTTCCAGAGCCCTCTCTCCCTCAGAGTCCAGAGGCCCCCTCTCTCC 800
QY 769 CTCAAACCAAGGTAGAGATCCCGAGCCCTCTCTCTCAGACCAAGAGTCCAGAGCC 828
    |||
Db 801 CTCAAACCAAGGTAGAGATCCCGAGCCCTCTCTCTCAGACCAAGAGTCCAGAGCC 860
QY 829 CCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 888
    |||
Db 861 CCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 920
QY 889 AGAC----- 893
    |||
Db 921 AGACCCCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 980
    |||
QY 893 ----- 893
    |||
Db 981 GGAATCCAGAGCCCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1040
QY 893 ----- 934
    |||
Db 1041 AGACCCAGAGAGTCCAGATCCCGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1100
QY 935 AACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 994
    |||
Db 1101 AACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1160
QY 995 AGGTNCAAGTGTCCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1054
    |||
Db 1161 AGGTNCAAGTGTCCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1220
QY 1055 CCTGTACAGAGTGTCCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1114
    |||
Db 1221 CCTGTACAGAGTGTCCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1280
QY 1115 TTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1167
    |||
Db 1281 TTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1333

```

```

FT /note- "Xaa - unknown"
PN M09820117-A1.
PD 14-MAY-1998.
PF 31-OCT-1997: U20051.
PR 05-NOV-1996: US-744026.
PI (INCY-) INCYTE PHARM INC.
PI Bandman O, Goli SR.
DR WPI: 98-286933/25.
DR P-PSDB: M60592.
PT New isolated prostate-specific kallikrein - used to develop products
PT for diagnosis and treatment of, e.g. prostate carcinoma or benign
PT hyperplasia
PS Claim 5: Fig 1A-C; 68BP; English.
CC This DNA encodes a human prostate-specific kallikrein (HPSK). A host cell
CC containing an expression vector comprising the HPSK nucleic acid sequence
CC can be used to produce the protein recombinantly. The HPSK products can
CC be used for the diagnosis of conditions or diseases associated with
CC expression of HPSK such as prostate carcinoma and benign prostate
CC hyperplasia. Agonists and antagonists which specifically bind to HPSK and
CC modulate its activity can be used for the preparation of treatment of
CC such conditions or diseases. The products can also be used for detection
CC and drug screening, especially for the detection of prostate-specific
CC kallikrein (PSK).
SQ Sequence 871 BP; 166 A; 260 C; 258 G; 184 T;

Query Match 61.8%; Score 721.4; DB 1; Length 871;
Best Local Similarity 98.2%; Pred. No. 3,6e-171;
Matches 725; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 2 CGCAGCCCTGCGAGCGGCGACGTGTCATGGAAGCAATTTGTCGCGGGCTCTGG 61
    |||
Db 134 CCAGCCCTGCGAGCGGCGACGTGTCATGGAAGCAATTTGTCGCGGGCTCTGG 193
QY 62 TGCATCCGAGTGGGTGCTCTGACCGCGACACTGTTTCCAGAACTCTTACCATCGGGC 121
    |||
Db 194 TGCATCCGAGTGGGTGCTCTGACCGCGACACTGTTTCCAGAACTCTTACCATCGGGC 253
QY 122 TGGGCTCTGCAAGTCTTGAAGCCGCGACCAAGAGCCAGAGTGTGAGAGCCAGCC 181
    |||
Db 254 TGGGCTCTGCAAGTCTTGAAGCCGCGACCAAGAGCCAGAGTGTGAGAGCCAGCC 313
QY 182 TCCCGTAGGAGCGACCGAGTACAGAGTCTTGTGCTGCTAAGCACTCATGCTATCA 241
    |||
Db 314 TCCCGTAGGAGCGACCGAGTACAGAGTCTTGTGCTGCTAAGCACTCATGCTATCA 373
QY 242 AGTTGAGCAATCCGTGCTGAGTGTGACACATCCGAGACATGAGATTGTTGCGAGT 301
    |||
Db 374 AGTTGAGCAATCCGTGCTGAGTGTGACACATCCGAGACATGAGATTGTTGCGAGT 433
QY 302 GCCCTACCGCGGGGAGACTCTTGCCTGCTNTCTGAGTGGGGTCTGTCGCGAGCGAA 361
    |||
Db 434 GCCCTACCGCGGGGAGACTCTTGCCTGCTNTCTGAGTGGGGTCTGTCGCGAGCGAA 493
QY 362 TGGCTACCGGCTGCACTGTGCTGAGTGTGCTGAGTGTGAGAGTGTGAGTAAGC 421
    |||
Db 494 TGGCTACCGGCTGCACTGTGCTGAGTGTGCTGAGTGTGAGAGTGTGAGTAAGC 553
QY 422 TCTATGACCCGCTGTACACACCCAGCATGTTCTGCGCGCGGAGGAGCAAGAGAG 481
    |||
Db 554 TCTATGACCCGCTGTACACACCCAGCATGTTCTGCGCGCGGAGGAGCAAGAGAG 613
QY 482 ACTCTGCAAGAGTGAAGTGTGGGGGCGCCCTGATCTGCAAGAGGATCTGAGAGGCTTG 541
    |||
Db 614 ACTCTGCAAGAGTGAAGTGTGGGGGCGCCCTGATCTGCAAGAGGATCTGAGAGGCTTG 673
QY 542 TGTCTTGGGAAAGCCCGGTGTGGCAACTTGGCGTCCAGAGTGTCTACCAACTCT 601
    |||
Db 674 TGTCTTGGGAAAGCCCGGTGTGGCAACTTGGCGTCCAGAGTGTCTACCAACTCT 733
QY 602 GCAAAATTCAGTGTGATGAGAAAAAGCTCCAGNCCAGTAACTCTGCGGAGTGGAA 661
    |||
Db 734 GCAAAATTCAGTGTGATGAGAAAAAGCTCCAGGCGCACTTAACCTTGGGAGTGGAA 793

```

QY 662 CCCATGAATTTGACCCCAATATACATCTGCGGANGAATTGAGAAATCTGTCCAG 721
 |||||
 DB 794 CCATGAATTTGACCCCAATATACATCTGCGGANGAATTGAGAAATATCTTCCAG 853
 |||||
 QY 722 CCCCTCTCTCTCCTGAGCC 739
 |||||
 DB 854 CCCCTCTCTCTCCTGAGCC 871
 |||||

RESULT 9

ID V58648 standard; cDNA: 1119 BP.
 AC V58648;
 DT 08-DEC-1998 (first entry)
 DE Prostate tumour specific gene clone DE14.
 KW Prostate tumour specific gene; human; prostate cancer; detection;
 therapy; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 34..528
 FT /*tag= a
 PN MO9837418-A2.
 PD 27-AUG-1998.
 PF 25-FEB-1998; U03690.
 PR 09-FEB-1998; US-904809.
 PR 25-FEB-1997; US-806596.
 PR 01-AUG-1997; US-904809.
 PA (CORI-) CORIXA CORP.
 PI Dillon DC, Xu J;
 DR WPI: 98-480805/41.
 DR P-PSDB: W69389.
 PT Novel human prostate specific tumour protein and fragments - useful
 for detecting and treating prostate cancers
 PS Claim 1: Page 116-117; 141pp; English.
 CC This sequence represents a human prostate tumour specific gene, and can
 be used in the method of the invention. The method is for detecting
 CC prostate cancer comprises contacting a biological sample with an agent
 CC able to bind an immunogenic portion of a prostate protein (such as
 CC encoded by this sequence). An antibody which binds to an immunogenic
 CC portion of the prostate protein, and the method can be used to detect,
 CC monitor progression of, or treat prostate cancers. The antibody may
 CC also be conjugated to a therapeutic agent for use in therapy of prostate
 CC cancers.
 SQ Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match 30.2%; Score 352.2; DB 1; Length 1119;
 Best Local Similarity 96.3%; Pred. No. 4.8e-79;
 Matches 360; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 CGAGGCCCTGGGAGGGGCGACGTGATGGAAGCAATTGTCGTGGGGCTCTGG 61
 |||||
 DB 8 CGAGGCCCTGGGAGGGGCGACGTGATGGAAGCAATTGTCGTGGGGCTCTGG 67
 |||||
 QY 62 TGCATCCGACAGTGGTGTCTGACCGGCACACTGTTCCAGAACTCTACACCATGGGC 121
 |||||
 DB 68 TGCATCCGACAGTGGTGTCTGACCGGCACACTGTTCCAGAACTCTACACCATGGGC 127
 |||||
 QY 122 TGGGCTGACAGTCTTTAGGCGGACCAAGAGCCAGAGAGCCAGATGGTGGAGCCAGCC 181
 |||||
 DB 128 TGGGCTGACAGTCTTTAGGCGGACCAAGAGCCAGAGAGCCAGATGGTGGAGCCAGCC 187
 |||||
 QY 182 TCTCCCTAGGCGGACCAAGATACAGAGACTTGTCTCCCTAAGACCTATCTATCA 241
 |||||
 DB 188 TCTCCCTAGGCGGACCAAGATACAGAGACTTGTCTCCCTAAGACCTATCTATCA 247
 |||||
 QY 242 AGTTGACGAAATCCGTGCTGAGTGTACACATCCGAGACATGAGATTGCTTCGAGT 301
 |||||
 DB 248 AGTTGACGAAATCCGTGCTGAGTGTACACATCCGAGACATGAGATTGCTTCGAGT 307
 |||||
 QY 302 GCCCTACCGGGGAAACTTGGCTGCTGCTGGGTGCTGGCGAAGCGGAGAA 361
 |||||

DB 308 GCCCTACCGGGGAAACTTGGCTGCTGCTGGGTGCTGGCGAAGCATGCTG 367
 |||||
 QY 362 TGCTACCGTGGTGG 375
 |||||
 DB 368 TGATTCGCAATCCAG 381
 |||||

RESULT 10

ID V61253 standard; cDNA: 1119 BP.
 AC V61253;
 DT 06-JAN-1999 (first entry)
 DE cDNA sequence of prostate tumour clone P703 splice variant DE14.
 KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
 OS Homo sapiens.
 PN MO9837093-A2.
 PD 27-AUG-1998.
 PF 25-FEB-1998; U03492.
 PR 09-FEB-1998; US-020956.
 PR 25-FEB-1997; US-806099.
 PR 01-AUG-1997; US-904804.
 PA (CORI-) CORIXA CORP.
 PI Dillon DC, Xu J;
 DR WPI: 98-609886/51.
 DR P-PSDB: W71873.
 PT Polypeptides comprising immunogenic portions of prostate proteins -
 used in a vaccine for the treatment of prostate cancer
 PS Claim 3: Page 108-109; 130pp; English.
 CC The present sequence is a new DNA which encodes an immunogenic portion
 CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
 CC can be used as a vaccine for the treatment of prostate cancer. The DNA
 CC was identified by analysis of a subtracted cDNA library obtained by
 CC subtracting a prostate tumour cDNA expression library with a normal
 CC tissue cDNA library.
 SQ Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match 30.2%; Score 352.2; DB 1; Length 1119;
 Best Local Similarity 96.3%; Pred. No. 4.8e-79;
 Matches 360; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 2 CGAGGCCCTGGGAGGGGCGACGTGATGGAAGCAATTGTCGTGGGGCTCTGG 61
 |||||
 DB 8 CGAGGCCCTGGGAGGGGCGACGTGATGGAAGCAATTGTCGTGGGGCTCTGG 67
 |||||
 QY 62 TGCATCCGACAGTGGTGTCTGACCGGCACACTGTTCCAGAACTCTACACCATGGGC 121
 |||||
 DB 68 TGCATCCGACAGTGGTGTCTGACCGGCACACTGTTCCAGAACTCTACACCATGGGC 127
 |||||
 QY 122 TGGGCTGACAGTCTTTAGGCGGACCAAGAGCCAGAGAGCCAGATGGTGGAGCCAGCC 181
 |||||
 DB 128 TGGGCTGACAGTCTTTAGGCGGACCAAGAGCCAGAGAGCCAGATGGTGGAGCCAGCC 187
 |||||
 QY 182 TCTCCCTAGGCGGACCAAGATACAGAGACTTGTCTCCCTAAGACCTATCTATCA 241
 |||||
 DB 188 TCTCCCTAGGCGGACCAAGATACAGAGACTTGTCTCCCTAAGACCTATCTATCA 247
 |||||
 QY 242 AGTTGACGAAATCCGTGCTGAGTGTACACATCCGAGACATGAGATTGCTTCGAGT 301
 |||||
 DB 248 AGTTGACGAAATCCGTGCTGAGTGTACACATCCGAGACATGAGATTGCTTCGAGT 307
 |||||
 QY 302 GCCCTACCGGGGAAACTTGGCTGCTGCTGGGTGCTGGCGAAGCGGAGAA 361
 |||||
 DB 308 GCCCTACCGGGGAAACTTGGCTGCTGCTGGGTGCTGGCGAAGCGATGCTG 367
 |||||
 QY 362 TGCTACCGTGGTGG 375
 |||||
 DB 368 TGATTCGCAATCCAG 381
 |||||

RESULT 11

ID X41114 standard; cDNA: 402 BP.

AC X41114; (first entry)
 DT 17-JUN-1999; EST EXP ID NO:58.
 DE Human secreted protein 5', EST: expressed sequence tag; diagnosis;
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW differentiation; hematopoiesis regulation; tissue growth regulation;
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
 OS Homo sapiens.
 PN M09306548-A2.
 PD 11-FEB-1999.
 PE 31-JUL-1998; IB1222.
 PR 01-AUG-1997; US-905135.
 PA (GIST) GENSET.
 PI Ducleit A, Dumas MLine Edwards J, Lacroix B;
 DR WPI: 99-153778/13.
 DR P-PSDB; Y12281.
 PT New nucleic acids encoding human secreted proteins - obtained from
 PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
 PT kidney, lung, umbilical cord, placenta and colon tissue
 PS Claim 1: Page 205, 824pp; English.
 CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
 CC secreted proteins, and encode the proteins given in Y12261 to Y12314,
 CC respectively. The proteins given represent the signal peptide and an
 CC N-terminal fragment of a secreted protein. The nucleic acid sequences
 CC can be used for producing secreted human gene products. They can also
 CC be used to develop products for diagnosis and therapy. The proteins
 CC obtained may have cytokine activity, cell proliferation/differentiation
 CC activity, hematopoiesis regulating activity, tissue growth regulating
 CC chemokinetic activity, anti-inflammatory activity, tumour inhibition activity/
 CC ligand activity, anti-inflammatory activity, tumour inhibition activity/
 CC and other activities. The products can be used in forensic, gene therapy
 CC and chromosome mapping procedures. The sequences can also be used for
 CC obtaining corresponding promoter sequences. The nucleic acids encoding
 CC the signal peptide can be used for directing extracellular secretion of
 CC a polypeptide or the insertion of a polypeptide into a membrane, or
 CC importing a polypeptide into a cell.
 SQ Sequence 402 BP; 80 A; 125 C; 121 G; 75 T;

Query Match 28.4%; Score 331.4; DB 1; Length 402;
 Best Local Similarity 96.2%; Pred. No. 5.5e-74;
 Matches 354; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 2 CGAGCCCTGCGAGCGGCTGTCATGAGAAACGAATTGTTGCTCGGGGCTCTGG 61
 |||||||
 DB 29 CGAGCCCTGCGAGCGGCTGTCATGAGAAACGAATTGTTGCTCGGGGCTCTGG 88
 QY 62 TGCATCCGCACTGGGCTGTCAGCGGCACTGTTTCCAGAA-----CTCTCT 109
 |||||||
 DB 89 TGCATCCGCACTGGGCTGTCAGCGGCACTGTTTCCAGAAAGTAGTKACAGACTCT 148
 QY 110 ACACCATCGGCTGGGCTGTCAGAGTCTTGAGGCGCAGCAAGAGCCAGGAGCCAGATGG 169
 |||||||
 DB 149 ACACCATCGGCTGGGCTGTCAGAGTCTTGAGGCGCAGCAAGAGCCAGGAGCCAGATGG 208
 QY 170 TGAAGGCCAGCTCTCGCTACGCGACCCAGAGTACACAGACTCTTGTCTTAAGACC 229
 |||||||
 DB 209 TGAAGGCCAGCTCTCGCTACGCGACCCAGAGTACACAGACTCTTGTCTTAAGACC 268
 QY 230 TCAGTCATCATCAAGTTGAGCAATCCGTTGCGAGTGCAGACATCCGGAGGAGTACAGCA 289
 |||||||
 DB 269 TCAGTCATCATCAAGTTGAGCAATCCGTTGCGAGTGCAGACATCCGGAGGAGTACAGCA 328
 QY 290 TTGCTTGCAGTGGCTTACCGCGGGAACCTTGGCTCTGTTTGGCTGGGGGCTCTGCTGG 349
 |||||||
 DB 329 TTGCTTGCAGTGGCTTACCGCGGGAACCTTGGCTCTGTTTGGCTGGGGGCTCTGCTGG 388
 QY 350 CGAAGGCG 357
 |||||||
 DB 389 CGAAGGCG 396

RESULT 12
 ID V58646
 AC V58646; standard; cDNA; 1459 BP.
 DT 08-DEC-1998 (first entry)
 DE Prostate tumour specific gene clone D56.
 KW Prostate tumour specific gene; human; prostate cancer; detection;
 KW therapy; ss.
 OS Homo sapiens.
 PN M09837418-A2.
 PD 27-AUG-1998.
 PE 25-FEB-1998; U03690.
 PR 09-FEB-1998; US-904809.
 PR 25-FEB-1997; US-806596.
 PR 01-AUG-1997; US-904809.
 PA (CORI-) CORIXA CORP.
 PI Dillon DC, Xu J;
 DR WPI: 98-480805/41.
 PT Novel human prostate specific tumour protein and fragments - useful
 PT for detecting and treating prostate cancers
 PS Claim 1: Page 114; 14pp; English.
 CC This sequence represents a human prostate tumour specific gene, and can
 CC be used in the method of the invention. The method is for detecting
 CC prostate cancer comprises contacting a biological sample with an agent
 CC able to bind an immunogenic portion of a prostate protein (such as
 CC encoded by this sequence). An antibody which binds to an immunogenic
 CC portion of the prostate protein, and the method can be used to detect,
 CC monitor progression of, or treat prostate cancers. The antibody may
 CC also be conjugated to a therapeutic agent for use in therapy of prostate
 CC cancers.
 SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Query Match 25.1%; Score 292.8; DB 1; Length 1459;
 Best Local Similarity 80.4%; Pred. No. 3.4e-64;
 Matches 390; Conservative 0; Mismatches 12; Indels 83; Gaps 1;

QY 92 ACTGTTCCAGAACTCTTACACCATCGGGCTGGGCTGCGACAGTCTTGGAGCGGACCAAG 151
 |||||||
 DB 25 AGTAGAGCGAGAGCTCTTACACCATCGGGCTGGGCTGCGACAGTCTTGGAGCGGACCAAG 84
 QY 152 AGCGAGGAGCGAGATGGTGGAGGCGAGCTCTCCGTAAGCGGACCCAGATACAAAGAC 211
 |||||||
 DB 85 AGCGAGGAGCGAGATGGTGGAGGCGAGCTCTCCGTAAGCGGACCCAGATACAAAGAC 144
 QY 212 TCTTGCTCGCTAAGCACTCATGCTCATCAAGTTGAGCAAGATCCGTTCCGAGTGTACA 271
 |||||||
 DB 145 CTTGCTCGCTAAGCACTCATGCTCATCAAGTTGAGCAAGATCCGTTCCGAGTGTACA 204
 QY 272 CCATCCGAGATCAGCATTTGCTGCGAGTCCCTACCGGGGAACTTTGCTCTGTT 331
 |||||||
 DB 205 CCATCCGAGATCAGCATTTGCTGCGAGTCCCTACCGGGGAACTTTGCTCTGTT 264
 QY 332 CTGGCTGGGGTCTGCGGCAAC----- 355
 |||||||
 DB 265 CTGGCTGGGGTCTGCGGCAACGTTGAGTCAAGGGTGTGTGCTCTTCAAGA 324
 QY 355 -----GGCAAGATGCTTAC 368
 |||||||
 DB 325 GGTCTCTGCGCACTGCGGGGGGCTGACCCAGAGCTGTGCTCCAGGCAAGATGCTTAC 384
 QY 369 CGTCTCACTGCTGAGACGTGCTGGTGTGTGAGGAGNGTGTGAGTAACTATGA 428
 |||||||
 DB 385 CGTCTCACTGCTGAGACGTGCTGGTGTGTGAGGAGNGTGTGAGTAACTATGA 444
 QY 429 CCGCTGTACACCCGAGATGTTCTGCGCGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 488
 |||||||
 DB 445 CCGCTGTACACCCGAGATGTTCTGCGCGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 504
 QY 489 CAAGC 493
 |||||||

PS Claim 3; Page 53-54; 130bp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 234 BP; 43 A; 68 C; 68 G; 55 T;

Query Match 19.68; Score 228.8; DB 1; Length 234;
Best Local Similarity 98.38; Pred. No. 1.9e-46;
Matches 230; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 203 ACAACAGACTCTTCGCTGCTAAGCAGACCTCATCATCAAGTGGAGCAATCCGATCCG 262
|||
Db 1 ACAACAGACCTTCTGCTGCTAAGCAGACCTCATCATCAAGTGGAGCAATCCGATCCG 60
QY 263 AGTCTGACACCATCCGAGCATCAGATTCCTTGCAGTCCCTACCGCGGGAACTCTT 322
|||
Db 61 AGTCTGACACCATCCGAGCATCAGATTCCTTGCAGTCCCTACCGCGGGAACTCTT 120
QY 323 GCTTCGINTCTGGCTGGGGTCTGCTGGGGAACGAGATCCTTACCGTCTGCACTGCG 382
|||
Db 121 GCTTCGINTCTGGCTGGGGTCTGCTGGGGAACGAGATCCTTACCGTCTGCACTGCG 180
QY 383 TGAACGTGTGCGGTGTGCTGAGGANGTCTGACAGTCAAGCTCTATGACCCGCTGT 436
|||
Db 181 TGAACGTGTGCGGTGTGCTGAGGANGTCTGACAGTCAAGCTCTATGACCCGCTGT 234

Search completed: September 28, 1999, 12:28:42
Job time: 5342 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:33:47 ; Search time 161.06 Seconds
(without alignments)
664.265 Million cell updates/sec

Title: US-09-030-606-175

Sequence: 1 GCGCAGCCCTGGCAGCGGC.....NTAAGAGAAGCGCAAAAAA 1167

Scoring table: IDENTITY_NUC

Searched: 176461 seqs, 45838279 residues

Database: Issued_Patents_NA.*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5C.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/5D.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS9.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	721.4	61.8	871	3	US-08-744-026-2	Sequence 2, Appl1
2	181.6	15.6	8174	1	US-07-914-281-5	Sequence 5, Appl1
3	181.6	15.6	8174	1	US-08-393-246-5	Sequence 5, Appl1
4	181.6	15.6	8174	1	US-08-525-058A-5	Sequence 5, Appl1
5	181.6	15.6	8174	5	PCT-US91-00899-3	Sequence 3, Appl1
6	154.6	13.2	1089	5	PCT-US96-04294-3	Sequence 1, Appl1
7	154.6	13.2	1089	5	PCT-US96-04294-3	Sequence 3, Appl1
8	153	13.1	986	4	US-08-557-146-1	Sequence 1, Appl1
9	142.2	12.2	4060	1	US-08-308-949A-1	Sequence 1, Appl1
10	140	12.0	732	3	US-08-361-395-2	Sequence 2, Appl1
11	125.4	10.7	833	4	US-08-790-137-2	Sequence 2, Appl1
12	95.8	8.2	937	3	US-08-684-862-11	Sequence 11, Appl1
13	85.8	7.4	1462	2	US-08-358-782D-14	Sequence 14, Appl1
14	84.2	7.2	992	2	US-08-358-782D-13	Sequence 13, Appl1
15	84.2	7.2	1729	4	US-08-844-024-1	Sequence 1, Appl1
16	81	6.9	832	5	PCT-US95-06157-5	Sequence 5, Appl1
17	81	6.9	760	5	PCT-US91-06157-7	Sequence 7, Appl1
18	81	6.9	766	5	PCT-US95-06157-9	Sequence 9, Appl1
19	80	6.9	1454	3	US-08-467-155A-2	Sequence 2, Appl1
20	80	6.9	1454	4	US-08-628-198-2	Sequence 2, Appl1
21	80	6.9	1454	5	PCT-US96-07343-2	Sequence 8, Appl1
22	75.4	6.5	699	3	US-08-738-413B-8	Sequence 8, Appl1
23	71.2	6.1	840	3	US-08-684-862-12	Sequence 12, Appl1
24	70.4	6.0	1333	3	US-08-684-862-9	Sequence 9, Appl1
25	66	5.7	988	3	US-08-684-862-10	Sequence 10, Appl1
26	65.6	5.6	1096	3	US-08-684-862-8	Sequence 8, Appl1
27	62.4	5.3	7218	2	US-08-232-463-14	Sequence 14, Appl1
28	57.8	5.0	821	3	US-08-650-129-1	Sequence 1, Appl1
29	57.8	5.0	821	3	US-08-650-129-2	Sequence 2, Appl1
30	57.8	5.0	866	3	US-08-650-129-3	Sequence 3, Appl1
31	54	4.6	2259	4	US-08-845-998-3	Sequence 5, Appl1
32	52.4	4.5	2218	4	US-08-845-998-5	Sequence 1, Appl1
33	51.4	4.4	200	3	US-08-931-981A-1	Sequence 1, Appl1
34	49	4.2	1554	3	US-08-469-486-1	Sequence 1, Appl1
35	47.2	4.0	8174	1	US-07-914-281-5	Sequence 5, Appl1
36	47.2	4.0	8174	1	US-08-393-246-5	Sequence 5, Appl1
37	47.2	4.0	8174	3	US-08-525-058A-5	Sequence 5, Appl1

ALIGNMENTS

38	47.2	4.0	8174	5	PCT-US91-00899-3	Sequence 3, Appl1
39	46.8	4.0	884	4	US-08-851-974-2	Sequence 2, Appl1
40	46	3.9	329	1	US-08-148-910-2	Sequence 2, Appl1
41	46	3.9	970	1	US-08-148-910-3	Sequence 3, Appl1
42	46	3.9	329	1	US-08-148-910-13	Sequence 13, Appl1
43	46	3.9	2033	1	US-08-148-910-14	Sequence 14, Appl1
44	46	3.9	329	2	US-08-448-937A-2	Sequence 2, Appl1
45	46	3.9	970	2	US-08-448-937A-3	Sequence 3, Appl1

RESULT 1

US-08-744-026-2
; Sequence 2, Application US/08744026
; Patent No. 5786148
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
; TITLE OF INVENTION: KALIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/744,026
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: Pf-0154 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 871 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: Consensus
; US-08-744-026-2

Query Match

61.8% Score 721.4; DB 3; Length 871;
Best Local Similarity 98.2%; Pred. No. 1.2e+184;
Matches 725; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY	2	CGCAGCCCTGGCAGCGGCACATGTCATGAAACGAATGTTCTCTCGGGCGCTCTGG	61
DB	134	CCAGCCCTGGCAGCGGCACATGTCATGAAACGAATGTTCTCTCGGGCGCTCTGG	193
OY	62	TGCATCCGAGTGGTGTCTGACGCCGACACTGTTCCAGAACTCTACACCATGGGC	121
DB	194	TGCATCCGAGTGGTGTCTGACGCCGACACTGTTCCAGAACTCTACACCATGGGC	253

QY	122	TGGGCTCTCACAGTCTTTAGGCGCCAGCAAGAGCCAGGGAGGCAAGTGTGTGAGGGCCACG	181
Db	254	TGGGCTCTCACAGTCTTTAGGCGCCAGCAAGAGCCAGGGAGGCAAGTGTGTGAGGGCCACG	313
QY	182	TCTCCGTACGGCACCAGCAGTACAAACAGACTCTTGCTGCTTAAGCAACCTCATGCTCATCA	241
Db	314	TCTCCGTACGGCACCAGCAGTACAAACAGACTCTTGCTGCTTAAGCAACCTCATGNTCATCA	373
QY	242	AGTTGGAGAGATCCGCTGTCCGAGTGTGACACCATCCGGAGCATAGCAATTGCTTCCGAGT	301
Db	374	AGTTGGAGAGATCCGCTGTCCGAGTGTGACACCATCCGGAGCATAGCAATTGNTTCCGAGT	433
QY	302	GCCCTACCCGCGGGGAACTCTTGCCCTCGTNTCTGGCGTGGGGTCTGCTGCGCAAGCGCAGAA	361
Db	434	GCCCTACCCGCGGGGAACTTTTGCCCTGTTGCGCTGGGGTCTGCTGCGCAAGCGCAGAA	493
QY	362	TGCTTACCGTGTGTCACATGTCGCGTGAACGTGTGCGGTGTCTTGAGGANGTCTGCAGTAAGC	421
Db	494	TGCTTACCGTGTGTCAGTGTGCGTGAACGTGTGCGGTGTCTTGAGGANGTCTGCAGTAAGC	553
QY	422	TCTATGACCCGCGTGTACACCCACCATGTTCTGGCGCGGGGAGGGAGGGAACCAAGAAAG	481
Db	554	TCTATGACCCGCGTGTACACCCACCATGTTCTGGCGCGGGGAGGGAGGGAACCAAGAAAG	613
QY	482	ACTCCTGCAAGGAGTACTCTGGGGGGCCCTCGATGTGCAACGGGTACTTGCAGGGCCTTG	541
Db	614	ACTCCTGCAAGGAGTACTCTGGGGGGCCCTCGATGTGCAACGGGTACTTGCAGGGCCTTG	673
QY	542	TGTCTTTGGAAAACCCCGCTGTGCGCAACTTGGCTGCCAGGTGTCTTACACCAACTCT	601
Db	674	TGTCTTTGGAAAACCCCGCTGTGCGCAAGTGGGTGCCAGGTGTCTTACACCAACTCT	733
QY	602	GCAATTTACAGAGTGTAGAGAAAACCGTCCACNCCAGTTAATCTGGGGAGCTGGGAA	661
Db	734	GCAATTTACAGTGTAGTGTAGAGAAAACCGTCCAGGCCAGTTAATCTTGGGGAGCTGGGAA	793
QY	662	CCCATGAATATGACCCCAATATCATCTGTGCGGAGANGAATTCAGGAATATCTGTTCCAG	721
Db	794	CCCATGAATATGACCCCAATATCATCTGTGCGGAGANGAATTCAGGAATATCTGTTCCAG	853
QY	722	CCCTCCTCCTCCCTCAGGCC	739
Db	854	CCCTCCTCCTCCCTCAGGCC	871

RESULT 2
 US-07-914-281-5/c
 Sequence 5, Application US/07914281
 Patent No. 532463
 GENERAL INFORMATION:
 APPLICANT: LOWE, JOHN B.
 TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
 OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
 TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
 TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
 ADDRESSEE: P.C.
 STREET: 1755 Jefferson Davis Highway, Fourth Floor
 CITY: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/914,281
 FILING DATE: 19920720
 CLASSIFICATION: 530

```

1  ATTORNEY/AGENT INFORMATION:
2  NAME: Lavallee, Jean-Paul M. P.
3  REGISTRATION NUMBER: 31,451
4  REFERENCE/DOCKET NUMBER: 2363-060-55
5  TELECOMMUNICATION INFORMATION:
6  TELEPHONE: (703)521-4500
7  TELEFAX: (703)466-2347
8  TELEX: 248855 OPAT UR
9  INFORMATION FOR SEQ ID NO: 5:
10 SEQUENCE CHARACTERISTICS:
11 LENGTH: 8174 base pairs
12 TYPE: NUCLEIC ACID
13 STRANDEDNESS: unknown
14 TOPOLOGY: unknown
15 MOLECULE TYPE: DNA (genomic)
16 ANTI-SENSE: NO
17 US-07-914-281-5

```

Query Match	15.6%	Score 181.6	DB 1	Length 8174
Best Local Similarity	77.3%	Pred. No. 1.2e-39		
Matches 269	Conservative	0	Mismatches 67	Indels 12
				Gaps 4
Qy 700	ATTCAGAAATATCTGTTC	CCACAGCCCCCTCTCTCCCTTAGAGCCACGAGATCGACAGCCCCCAGC	759	
Db 1025	ACTTAGGAGTCCAGGCTCTCCGGCCCCCTCTCTCTTAGACACGAGATCCAGCCCCCTTGC	966		
Qy 760	CCCTCTCTCTCTCAAAACAGG-GTAGAGATCCACAGCCCCCTCTCTCTAGACCCAGAG	818		
Db 965	CCCTCTCTCTCTAGACCCAGAGGTCCAGAGACCCAGCCCCCTCTCTCTACACCCAGAG	906		
Qy 819	TCACAGCCCCCAGCCCCCTCTCTCTAGACCCAGAGATCCAG-----CCCTCTCTC	870		
Db 905	TCACAA-TCCTTAGCCCCCTACTCTCTCTAGACCCAGAGATCCAGACCAAGCTCTCTCTC	847		
Qy 871	CNTCAGACGAGAGATCCAGACCCCCAGACCCCTNTCTCGTAGACCCAGAGGGTCCAGGC	930		
Db 846	CCTCAGACCCAGAGCCCAAGTGTCCACGCCCTCTCTCTTAGATCCAGAGGTACAGG-787			
Qy 931	CCCCAACCCCTCTCTCTAGAGTCCAGAGGTCCAGAGCCCCCAACCCCTGTCTCCAGAGC	990		
Db 787	-CCGAGACCTCTCTCTCTAGACCCAGAGATCCAGAGCCCCCAACCCCTCTCTCTAGAGC	729		
Qy 991	CCAGAGTNCAGAGTCCAGCCCCCTCTCTCTCTAGACCCAGAGGGTCCAA	1038		
Db 728	CCAGAGTTCAGAGCCCAAGCCCTCTCTCTCTAGACACAGAAAGGCTTA	681		

RESULT 3
 US-08-393-246-5/c
 Sequence 5, Application US/08393246
 Patent No. 5595900
 GENERAL INFORMATION:
 APPLICANT: LOWE, JOHN B.
 TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
 OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
 TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
 TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
 ADDRESSEE: P.C.
 STREET: 1755 Jefferson Davis Highway, Fourth Floor
 CITY: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:

```

1  APPLICATION NUMBER:  US/08/393,246
2  FILING DATE:
3  CLASSIFICATION:  530
4  PRIOR APPLICATION DATA:
5  APPLICATION NUMBER:  US 08/220,433
6  FILING DATE:  30-MAR-1994
7  APPLICATION NUMBER:  US 07/914,281
8  FILING DATE:  20-JUL-1992
9  ATTORNEY/AGENT INFORMATION:
10 NAME:  Lavalleye, Jean-Paul M. P.
11 REGISTRATION NUMBER:  31,451
12 REFERENCE/DOCKET NUMBER:  2363-060-55
13 TELECOMMUNICATION INFORMATION:
14 TELEPHONE:  (703)521-4500
15 TELEFAX:  (703)486-2347
16 TELEX:  248855 OPAT UR
17 INFORMATION FOR SEQ ID NO:  5:
18 SEQUENCE CHARACTERISTICS:
19 LENGTH:  8174 base pairs
20 TYPE:  nucleic acid
21 STRANDEDNESS:  unknown
22 TOPOLOGY:  unknown
23 MOLECULE TYPE:  DNA (genomic)
24 ANTI-SENSE:  NO
25 US-08-393-246-5

```

Query March 15.6% Score 181.6; DB 1; Length 8174;
Best Local Similarity 77.3%; Pred. No. 1.2e-39;
Matches 269; Conservative 0; Mismatches 67; Indels 12; Gaps 4;

DY 700 ATTGAGAAATCTGTTCACAGCCCTCCTCCTCAGGCCCGAGAGTCCAGGCCCCCAGC 759
| | | | | | | | | | | | | | | | | | | | | |
DB 1025 ACTTAGAGATCCAGGCTCCGGCCCCCTCCTCTCAGACCAGAGATCCAGGCCCTCAGC 966
| | | | | | | | | | | | | | | | | | | | | |
QY 760 CCGCTCTCCCTCAACAAG-GTACAGATCCCAAGCCCCCTCCTCCTCAGACCCAGAG 818
| | | | | | | | | | | | | | | | | | | | | |
DB 965 CCCTCTCTCTCAGACCCAGAGGTCCAGAGCCCCCAGCCCTCTCTCTCACACCCACAG 906
| | | | | | | | | | | | | | | | | | | | | |
QY 819 TCCAGACCCCGAGCCCTCCTCCTCAGACCCAGAGAGTCCAG-----CCGCTCTC 870
| | | | | | | | | | | | | | | | | | | | | |
DB 905 TCACAGA-TCCCTAGCCCCCTACTCTCCTCAGACCCAGAGATCCAGACCAAAGCTCTCTC 847
| | | | | | | | | | | | | | | | | | | | | |
QY 871 CNTCAGACGAGAGTCCAGACCCCGAGCCCTCCTCCTCAGACCCAGAGGGTGACAGC 930
| | | | | | | | | | | | | | | | | | | | | |
DB 846 CCTCAGACCCAGAGGCCAAGTTCCCGAGCCCTCCTCCTCAGATCCAGAGAGTACAG - 787
| | | | | | | | | | | | | | | | | | | | | |
QY 931 CCCCAACCCCTCTCCTCCTCAGAGTCCAGAGTCCAGGCCCCCAACCCCTGTGTTCCCAAGC 990
| | | | | | | | | | | | | | | | | | | | | |
DB 787 -CCAGACCCCTCTCCTCTCAGACCCAGAGAGTCCAGGCCCCCGACCCCTCTCTCTCAGAC 729
| | | | | | | | | | | | | | | | | | | | | |
QY 991 CCAGAGTNCAGAGTCCAGGCCCTCCTCCTCAGACCCAGAGGGTGCAA 1038
| | | | | | | | | | | | | | | | | | | | | |
DB 728 CCAGAGTCCAGAGGCCCGAGCCCTCTCCTCCTCAGACAGAAAGGCTTA 681

RESULT 4
US-08-525-058A-5/c
Sequence 5, Application US/08525058A
Patent No. 5770420
GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS OR AS FREE MOLECULES, AND FOR THE ISOLATION
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLOH, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P.C.
STREET: 1755 Jefferson Davis Highway, Fourth floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

```

1      ZIP: 22202
2      COMPUTER READABLE FORM:
3      MEDIUM TYPE: Floppy disk
4      COMPUTER: IBM PC compatible
5      OPERATING SYSTEM: PC-DOS/MS-DOS
6      SOFTWARE: Patent Release #1.0, Version #1.25
7      CURRENT APPLICATION DATA:
8      APPLICATION NUMBER: US/08/525,058A
9      FILING DATE:
10     CLASSIFICATION: 435
11     ATTORNEY/AGENT INFORMATION:
12     NAME: Lavalleye, Jean-Paul M. P.
13     REGISTRATION NUMBER: 31,451
14     REFERENCE/DOCKET NUMBER: 2363-060-55
15     TELECOMMUNICATION INFORMATION:
16     TELEPHONE: (703)521-4500
17     TELEFAX: (703)486-2347
18     TELEX: 24885 OPAT UR
19     INFORMATION FOR SEQ ID NO: 5:
20     SEQUENCE CHARACTERISTICS:
21     LENGTH: 8174 base pairs
22     TYPE: nucleic acid
23     STRANDEDNESS: double
24     TOPOLOGY: linear
25     MOLECULE TYPE: DNA (genomic)
26     ANTI-SENSE: NO
27     US-08-525-058A-5

```

```

Query Match      15.6%: Score 181.6; DB 3; Length 8174;
Best Local Similarity 77.3%: Pred. No. 1.2e-39;
Matches 269; Conservative 0; Mismatches 67; Indels 12; Gaps 4;

OY 700 ATTCAGGAAATATCTGTTCCAGGCCCTCTCTCCCTCAGGCCCCAGAGTCCAGGCCCCAGC 759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1025 ACTTAGAGGTCCAGAGCTCCCGGCCCTCTCTCTCCACAGCAGAGTCCAAAGCCCTTC 966
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 760 CCGCTCCCTCCCTCAACCAAGG-GTACAGATCCCGAGGCCCTCTCTCCCTCAGACCCAGAG 818
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 965 CCGCTCTTCTCTCAAGCCAGAGTCCAGAGCCCAAGCCCTCTCTCTCAACCCAGAG 906
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 819 TCAGAGCCCCCAGCCCCCTCTCTCTCAAGCCAGAGTCCAG-----CCCTCTCTC 870
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 905 TCCAGA-TCCCTACCCCTTACTCTCTCAGACCCAGAGATCCAGACCAAGCTCCCTCTC 847
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 871 CNTAGAGCAGAGGTCTCAGACCCCGAGCCGCTCTCTCTCCGTCAGACCCAGGAGTCCAGG 930
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 846 CCTCAGACCCAGAGCCCAAGTCCCAAGCCCTCTCTCTCTCAATCAGAGATCAGAG- 787
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 931 CCCCAGACCCCTCTCTCTCAAGTCTCAGAGTCCAGAGTCCAAAGCCCCCAAGCCCTCTCTCC 990
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 787 -CCAGAGCCCTCTCTCTCTCAGACCCAGAGAGCGGCCCCCAAGCCCTCTCTCTCAGAC 729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 991 CCAGAGTCCAGAGTCCCAAGCCCTCTCTCTCTCTCAGACACCAAGAGTCCAA 1038
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 728 CCAGAGTCCAGAGCCCAAGCCCTCTCTCTCTCAGACACAGAGAGCTTA 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
PCT-US91-00899-3/C
: Sequence 3, Application PC/TUS9100899
:
: GENERAL INFORMATION:
:
: APPLICANT: Lowe, John B.
:
: TITLE OF INVENTION: Method and Products For the Synthesis of
: TITLE OF INVENTION: Oligosaccharide Structures on Glycoproteins, Glycolipids,
: TITLE OF INVENTION: or as Free Molecules, and For the Isolation of Cloned
: TITLE OF INVENTION: Genetic Sequences That Determine These Structures
: NUMBER OF SEQUENCES: 16
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: OBION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
: ADDRESSEE: P.C.
: STREET: 1755 Jefferson Davis Highway, Suite 400
: CITY: Arlington
:

```

STATE: Virginia
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/00899
FILING DATE: 19910214
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye Ph.D., Jean-Paul
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-021-55 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-5940
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 8174 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Blood
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4686..5780
OTHER INFORMATION: /label=mat_peptide
PCT-US91-00899-3

TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
 TITLE OF INVENTION: RELATED NUCLEIC ACIDS
 NUMBER OF SEQUENCES: 3
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Eli Lilly and Company
 STREET: Lilly Corporate Center
 CITY: Indianapolis
 STATE: Indiana
 COUNTRY: United States of America
 ZIP: 46285
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US96/04294
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/416,257
 FILING DATE: 04-APR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Bialock, Donna K.
 REGISTRATION NUMBER: 38,082
 REFERENCE/DOCKET NUMBER: X9239
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 317-277-1090
 TELEFAX: 317-276-3861
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1089 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA
 PCT-US96-04294-1

```

Query Match          15.6%; Score 181.6; DB 5; Length 8174;
Best Local Similarity 77.38; Pred. No. 1,2e-39;
Matches 269; Conservative 0; Mismatches 67; Indels 12; Gaps

QY 700 ATTCAGGATATCTCTTCCACGCCCTCCTCCTCCTCAGGCCCAAGAGATCCAGGCCCCCAGC 759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1025 ACTTAGGATCCAGGCTCCGGGCCCTCCTCTTCCTCAGACCAGAGATCCAAAGCCCCCTCAG 966
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 760 CCCTCCTCCTCAAAACCAAG-GTACAGATCCCCAGCCCCCTCCTCCTCAGACCAGAG 818
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 965 CCCTCCTCCTCAGAGCCAGGAGTCCAGGACCCCAACCCCTCCTCCTCAGACCCACAGAG 906
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 819 TCCAGACCCCCAGGCCCTCCTCCTCCTCCTCAGCCAGGAGTCCAG-----CCCCCTC 870
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 905 TCCGAA-TCCTAGCCCCCTACTCCTCCTCAGACCAGAGATCCAGACCAAAAGCTCCTCCTC 847
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 871 CMTGAGCGAGAGTCCAGACCCCCCAAGCCCTCCTCCTCAGAGCCAGGGGTGCAGGC 930
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 846 CCTCAGACCCAGAGACCCCAAGTTCGCCACCCCTCCTCCTCAGATCCAGAGTACAGG- 787
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 931 CCCCAACCCCTCCTCCTCCTCAGAGTCCAGAGTCCAGCCCCCAACCCCTCCTCCTCCAGAC 990
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 787 -CCGAGACCTCCTCCTCAGACCAGAGTCCAGGCCCCCCCAACCCCTCCTCCTCCTCAGAC 729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 991 CCAGAGGTCAGAGTCCAGACCCCTCCTCCTCCTCAGACCAGCCAGCGGTCCAA 1038
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 728 CCAGAGTCCAGAGCCGCCCAAGCCCTCCTCCTCAGACAGACAGAAAGGCTTA 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT      6
PCT-US96-04294-1/c
; Sequence 1, Application PC/TUS9604294
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
;

```

Query Match	13.2%	Score 154.6	DB 5	Length 1089
Best Local Similarity	55.1%	Pred. No. 9.7e-33		
Matches 351	Conservative 0	Mismatches 271	Indels 15	Gaps 2
QY 2	CGCAGCCCTGGCAGCGCGCACATGTCATATGGAAGAAATTTCTGCTCGCGGCTCTGG 61			
Db 855	CCACCCCATGGCAGATGGCTGCTGCTCATAGGGGAATACACTCCACTGTGGGAGGGCTCTGG 796			
QY 62	TGCATCCGCAGTGTGGTGTCTGTGACGCCGACACTGTTTCCAGACTCCTACACCATGGGCG 121			
Db 795	TCATATGAGCCCTGGGTGCTCATCTGGCGCCCACTGTCAGATGAAATGATACACGTCGACC 736			
QY 122	TGGGCCCTGCACATCTGTGAGGCGCGAACAAAGCCAGGAGGACAGATGTGGAGGCCAGCC 181			
Db 735	TGGGCACTGATACGCTGGGCGCACAGAGAG-----CTCAGAGATCAAGGCCCTCGA 865			
QY 182	TCCTCCGACGGCACCAGAGTACAAACAGACTTCTGCTGCTACAGACCTATGCTATCA 241			
Db 684	AGTCATTCGGCCACCCCGGCTACTCCACACAGACCCATGTTAATGACTCATATGCTGTGA 625			
QY 242	AGTTGGACGAATCCGTGTCCGAGCTGACACCATCCGGAGCATAGCATTTGCTTGCAGT 301			
Db 624	AGCTCATATACCAGGCGCAGAGCTGTCATCCATGATGTAAGAAAGTACAGGCTGCCTCCCGCT 565			
QY 302	GCCCTACCGGGGGAGACTTGTGCTCGTNTCTGGGTGGGGTCTCTGCGGAC-----G 355			
Db 564	GCGACCCCCCTGGAAACACACTGTACTGTCTCTCGGGTGGGGCACTACACAGAGCCCATG 505			
QY 356	GCAGATAGCTACCGTGTCTGCACTGCGTGAAGAGTGTCCGTGGTGTCTGAGGANGTCTGCA 415			
Db 504	TGACCTTTCCTCTGACCTCATATGTGATGATGATCAACCTCATCTCCGCCAGGACGTGCA 445			
QY 416	GTAAGCTCTATGACCCGCTGTACCAACCCAGAGATTTCTGCGCGGCGAGGCGAAGACC 475			

Db	444	CGAAGTTTACCAAGACTTACTGAAAAATTCANGTGCTGGCTGGGATCCCGACTCCA	385
Qy	476	AGAAAGGACTCCTGCACAGCGTACTCTCTGGGGGGGCCCTGATCTGCACAGGGTACTTGAGG	535
Db	384	AGAAAAAGCCTTCGAATGTGTACTCAGAGGGGAGCCGTTGGTGTGCAGAGTACCCCTTGCAAG	325
Qy	536	GCCTGTGTCTTTGGGAAAAAGCCCCGTGTGGCCAACTTGGCGTGCAGGTGTACACCA	535
Db	324	GTCGTGGTCTCTGGGGAACCTTCCCTTGGCCCAACCAATGACCCAGAGGTACTACCTC	265
Qy	596	ACCTCTGCAATTTACTGTAGTGGATAGAAAAACCGT	632
Db	264	AAGTCTGCAGTTTACCAACGTGAGTAAATGACACCAT	228

```

RESULT 7
PCT-US96-04294-3/C
: Sequence 3, Application PC/TUS9604294
GENERAL INFORMATION:
APPLICANT: Dixon, Eric P.
APPLICANT: Johnstone, Edward M.
APPLICANT: Little, Sheila P.
TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
TITLE OF INVENTION: RELATED NUCLEIC ACIDS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/04294
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/416,257
FILING DATE: 04-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Bialock, Donna K.
REGISTRATION NUMBER: 38,082
REFERENCE/DOCKET NUMBER: X9239
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-277-1090
TELEFAX: 317-276-3861
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1089 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: mRNA
PCT-US96-04294-3

Query Match 13.2% Score 154.6; DB 5; Length 1089;
Best Local Similarity 55.1%; Pred. No. 9,7e-33;
Matches 351; Conservative 0; Mismatches 271; Indels 15; Gaps

OY 2 CGCAGCCCTGGCAGCGGCGACATGCTATGGAAGAATTGTTCTGCTGGGCGTCTGG 61
Db 855 CCCACCACATGGCAGATGGCCCTGCTCTCATGGCATTGCACTTCACGCGAGGGCTCTGG 796
795 TCATATGACCGCTGGGTGCTACTCTCCGCCCACTCTCAAGATGAATGAGTACACCGTGCAAC 736

```

QY	122	TGGGCGCTGCAAGCTCTTGAGGCGGACCAAGAAGCCAGGACCAGATGATGATGAGGCCAGC	181
Db	735	TGGGCACTGATAGCTCTGGGCGACAGGAGG-----CTCAGAGGATTAAGGCGCTCGA	685
QY	182	TTCTCGTAGCGGACCCAGAGTACAAACGACTCTTGCTCGCTAAAGAGACTCATGCTATCA	241
Db	684	AGTCATTCCGCGACACCCGGGCTACTCCACACAGAACCCATGTTAAATGACTCATGCTGTGA	625
QY	242	AGTTGACGAATCCGTGTCCGAGTCTGACACCAATCCGAGACATCAGCATTTGTTGCGAGT	301
Db	624	AGTCATATAGCAGAGGCCAGGCGTGTCAATCCATGGTGAAGAAGTCAAGGCTCCCTCCCGCT	565
QY	302	GCCCTACCGCGGGGAACTCTTGCCCTCGTNTCTGGCTGGGGTCTGTGGCGAAC-----G	355
Db	564	GCGAACCCCTCGGAACCACTGTACTGTCTCCGGCTGGGGCACTACCAAGAGCCAGATG	505
QY	356	GCAAAATGCCATACCGTCTGCACTGCGTGAACGTGTGTGGTGTCTGAGAGATGTGCA	415
Db	504	TGACCTTTCCCTTGACCTCATGTGCGTGAGATCAAGTCAATCTCCCCAGAGACTGCA	445
QY	416	GTAAGCTCTATGAAACCGGCTGACACACCCAGCAATGTTCTCGCGCGGGGAGGCAAGAC	475
Db	444	GGAAGGTTTACAAAGACTTACTGAAAAATTCCATGCTGTGCGCTGTGGCATCCCCGACTCA	385
QY	476	AGAAGACTCCTGCACAGGTGACTCTGGGGGGGCCCTGATCTGCAACGGGTACTTGCAGG	535
Db	384	AGAAAAAGCGCTGCAAATGTGTACTCAGGGGAGCCGTTGGTGTGACAGAGTAACTCTGAAG	325
QY	536	GCCTTGTGTCTTTGGAAAAAGCCCGGTGTGGCCAATTTGGCGTGGCCAAGTGTCTACCA	595
Db	324	GTCGTGGTGTCTCGGGAACCTTCCCTTTCGCGCCAAACCAATGACCAAGAGTCTACACTC	265
QY	596	ACCTGTCAAAATTCACGTAGTGGATGGATGAGAAAAACCG	632
Db	264	AAGTCTGCAAGTTCCACCAAGTGGATTAATACACCAT	228

RESULT 8
US-08-557-146-1
Sequence 1, Application US/08557146
Patent No. 5834290
GENERAL INFORMATION:
APPLICANT: Egelrud, Torbjorn
APPLICANT: Hansson, Lennart
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
TITLE OF INVENTION: Enzyme (SCCE)
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESSES:
ADDRESSEE: White & Case, Patent Department
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2787
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557.146
FILING DATE: 14-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Sterner, Richard J.
REGISTRATION NUMBER: 35,372
REFERENCE/DOCKET NUMBER: 1103326-181
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 819-8783
TELEFAX: (212) 354-8113
INFORMATION FOR SEO ID NO.: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 986 base pairs

```

;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 25..786
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 25..90
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 112..783
;
US-08-557-146-1

```

```

Query Match          13.1%; Score 153; DB 4; Length 986;
Best Local Similarity 54.9%; Pred. No. 2.5e-32;
Matches 350; Conservative 0; Mismatches 272; Indels 15; Gaps 2;

```

```

QY 2 CGCAGCCCTGGCAGCGGCGCATGTGATGAAACGAAATGTTCTGCTCGGCGTCTTG 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 143 CCACACCATGGCAGGTGGGCGCTGCTCAGTGCATACGCTCCACTGGGAGGGCTCTG 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 TGCATCCGCACTGGTGTCTCAGCCGACACTGTTCCAGAACTCTACACATGGGGC 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 203 TCATATAGGCGCTGGTGTCTCTACTGCGCCCATGCAAGATGAATGATACCGTGCACC 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 122 TGGGCTGCACACTCTTGTAGGCGGACAGGCGAGGACGATGATGGTGGAGGCCAGCC 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 263 TGGGCGATGATAGCTGGGCGAGAGAG-----CTCAGAGGATAAAGGCTCCGA 313
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 182 TCTCCGTACGCGACCCAGAGTACAGACTCTTGTCTGCTTACAGACTGCTATCA 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 314 AGTATTCGCGCACCCCGGCTACTCCACACAGACCATGTTAATGACCTATGCTGTGA 373
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 242 AGTTGACGAAATCCGTGTGCGAGTCTGACACCATCCGGAGCATCAGATTGCTTGCAGT 301
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 374 AGTCTAATAGCCAGGCGAGGCTGTATCCATGTTGAAGAAAGTCAGGCTGCCCTCCGCT 433
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 302 GCCCTACCGGCGGGAATCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 355
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 434 GCAACCCCTGGACACACTGTACTGTCTCCGCTGGGCGACTTACCACGACCCAGATG 493
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 356 GCAGAAATGCTACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 415
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 494 TGACCTTTCCTGTGACCTGATGTGGTGTGATGCAAGCTCATCTCCCCCAGAGCTGCA 553
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 416 GTAACTCTATGACCGGCTTACCAACCCAGCATGTTCTCCGCGGCGGAGGCGCAAGACC 475
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 554 CGAAGGTTTACAAGGACTTCTGAAATTCATGCTGTGCTGCTGCTGCTGCTGCTGCTGCA 613
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 476 AGAAGGACTCTGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGAG 535
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 614 AGAAAAAGGCTTCAATGTGACTCAGGGGAGCGGTTGGTGTGAGAGGATACCTGGAAG 673
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 536 GCCTTGTGCTTTCGAAAGGCGCGGCTGTGCGCAACTTGGCTGCTGCTGCTGCTGCTGCTGCA 595
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 674 GTGTGTGTCTGCTGCGGAACTTTCCTTGGGCGCAACCCATGAGCCAGAGTCTACATC 733
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 596 ACCTGTGCAAAATTCAGTGAATGATGAGAAAACCT 632
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 734 AAGTGTGCAAGTTCACCAAGTGAATGATGACACCAT 770
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 9
; Sequence 1, Application US/08308949A
; Patent No., 5580703

```

```

;
; GENERAL INFORMATION:
; APPLICANT: Rotin, Robert M.
; APPLICANT: Berns, Kenneth I.
; APPLICANT: Linden, Ralph M.
; TITLE OF INVENTION: Human Adeno-Associated Virus Integration
; TITLE OF INVENTION: Site DNA and uses thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/308,949A
; FILING DATE: September 20, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/947,127
; FILING DATE: September 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Alice O.
; REGISTRATION NUMBER: 33,542
; REFERENCE/DOCKET NUMBER: ACC92-10P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-9540
; TELEFAX: 617-861-6240
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-08-308-949A-1

```

```

Query Match          12.2%; Score 142.2; DB 1; Length 4060;
Best Local Similarity 71.6%; Pred. No. 3.4e-29;
Matches 240; Conservative 0; Mismatches 81; Indels 14; Gaps 4;

```

```

QY 718 CCAGCCCTCTCTCCCTCAGAGCCAGAGAGTCCAGGCGCCCGCTCTCTCTCAACCA 777
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4000 CCAACCCCTCCCAATCAACCAAGAGGCGGCGCCAGCCCTTCCGCTCAGATGAA 3941
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 778 AGGATCAATCCCAAGCCCTCTCTCTCAGAGCCAGAGAGTCCAGAGCCCGCCCT 837
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3940 GGAATCCAGGCGCCAGGCTCTCTCTCTCAGAGCCAGAGGAGTCCAG--GCCAGAGCCG 3884
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 838 CMTCTCAGACCCAGAGTCCAG-----CCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 889
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3883 CTTCTTAAGACCCAGAAATCCAGAGGCGCCCGCCAGCCCTCTCTCTCTCTCTCTCTCTCA 3824
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 890 GACCCCGCAGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 949
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3823 G--GCCCGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3766
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 950 AGAGTGAAGTTCAGAGCCCGCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1009
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3765 AGAGCCAGAGAGTCC--AGGCGCCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3707
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1010 CCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1044
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3706 CCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3672
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 10

```

```

US-08-361-395-2
: Sequence 2, Application US/08361395
: Patent No. 5733768
:
: GENERAL INFORMATION:
: APPLICANT: Dixon, Eric P.
: APPLICANT: Johnstone, Edward M.
: APPLICANT: Little, Sheila P.
: APPLICANT: No. 5733768-1's, Franklin H.
: TITLE OF INVENTION: PROTEASE AND RELATED DNA COMPOUNDS
: NUMBER OF SEQUENCES: 8
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Eli Lilly and Company
: STREET: Lilly Corporate Center/Patent Division
: CITY: Indianapolis
:
: STATE: Indiana
:
: COUNTRY: United States
:
: ZIP: 46285
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/361,395
: FILING DATE: 22-DEC-1994
: CLASSIFICATION: 530
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Blalock, Donna K.
: REGISTRATION NUMBER: 38,082
: REFERENCE/DOCKET NUMBER: X8350A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-277-1090
: TELEFAX: 317-276-3861
:
: INFORMATION FOR SEQ ID NO: 2:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 732 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: linear
:
: MOLECULE TYPE: CDNA
:
US-08-361-395-2

```

Query Match	Similarity	12.0%	Score 140;	DB 3;	Length 732;
Best Local	Similarity	52.1%	Pred. No. 6.9e-29;		
Matches	333;	Conservative	0;	Mismatches 303;	Indels 3; Gaps 1.
0Y	2	CGCACCCCTGGAGACGGCAGCATGTGATGGAACGAAATGTTGTCGTGGGCGTCTCG	61		
Db	95	CTACACCCCTACCAACAGTGGCCCTCTACACCTCGGGGCACCTTGCTCTGTGGTGGGCTCTTA	154		
0Y	62	TGCATCCGAGTGGGTGCTGTGACGCCGACACTGTTTCCAGAACTCCTACACATGGGCG	121		
Db	155	TCCATCCACTGTGGGTCACAGCTGCACGCTGCCACCTGCAAAAAACGAAATCTTCAGGTCTTC	214		
0Y	122	TGGGCTTGACAGTCTTGAGGCCGACCAAGACCCAGGACCATGTTGAGGCCAGCC	181		
Db	215	TGGGGAACATTAACCTTCGGCAAGGAAAGTTCCAGAGCAAGATTCCTGTTG--TTC	271		
0Y	182	TCTCGTACGGGACCCAGATGACAAAGACTTTCGTGCTAACGACCTCATGTCATATA	241		
Db	272	GGGCTGTGATCCACCCTGTGACTGTGATGCGCCGACCATACCAAGGACATCAATGCTGTTC	331		
0Y	242	AGTTGAGCAATCCCTGTTCGAGTGTGACACCATCCGAGCAATCAAGATTCGTCAGT	301		
Db	332	GGCTGGCAACCCCAACGCAAACTCTGAACTATCCAGCCCTTCCCTCGGAGAGGAACT	391		
0Y	302	GCCCTACGCGGGGAACTTTCCTCTGTTTCGCTGGGGTCTGCTGGCAACGGCAGAA	361		
Db	392	GCTACGCCAACACACACACTGCCATCTCTGGCTGGGGCAAGACAGCAGATGGTATT	451		
0Y	362	TGCCATACGCTGCTGCACCTCCGTGAAGCTCGGTGCTCTGAGGANGTCTGCAGTAAC	421		

Db	452	TCCTGCAGACCATCCAGTGTGCATATACATCCACCTGGTGTCTCCCTGGAGAGATGTGAGCATG	5111
Qy	422	TCGTATGACCCCGCTGTACCACCCCAAGCATGTTCCTGCGCCGGCGGAGGCCAAGACAGCAAG	481
Db	512	CCCTACCCCTGGCCCAATATACCCCGAAGCATGTGTGTGTGTGGGATGAGAACTACGGGGAAG	5711
Qy	482	ACTCTGCACAGGTGACACTGTGGGGGGGCCCGATCTGCACAGGGTACTTGCAGGGCCCTTG	5411
Db	572	ATTCTCTGCAGAGGTGATTCTGGGGGGTCCGCTGGTATGTGGAGACACACTCTCCAGGCCCTTG	6311
Qy	542	TGCTTTTCGAAAAACCCCGTGTGGCCAACTTGGCGTGGCAGGTGTCTACCAACCACTCT	601
Db	632	TGTGATGGGGTAAACATCCCCCTGTGTGATCAACGAAGGAAAGCCAGGAGTCTACCAACAACCTCT	6911
Qy	602	GCAATTCACGTAGTGGATGAGAAAAACGGTCAGNACA	640
Db	692	GCAGTACACGAACTGGATCCAAAAAACCATTCAGGCCCA	730

```

11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1
```

[illegible]

DB 190 TNCACCCAGATGTTCTGACAGAGCCAGCTGCTNNAAGCCCGCTACATAGTTCACC 249
QY 122 TGGGCTGACATCTTTGAGCCGACAGAGCCAGAGGAGCCAGATGTTGAGAGCCAGCC 181
DB 250 TGGGGAGACACACCTCCAGAGAGAGAGGCTGTAG--CAGACCCGAGACGACTG 306
QY 182 TCCTCGTACGAGCCAGAGATACAGACTCTTGGCTGCTAGCAGC-----230
DB 307 AGTCCTTCCCGACCCCGGTTACACAGAGCTCCCGACAAAGACACCGCAATGACA 366
QY 230 TCATGCTATCAAGTTGAGAGAAATCCGTTCCAGTCTGACACCATCCGAGATGACA 289
DB 367 TCATGCTGAGAGATGAGATCCGACAGTCCATCACTGGGCTGTGCGACCCCTACCC 426
QY 290 TTGCTTCGAGTCCCTACCGGGGGAACCTTGCTGCTGCTGCTGCTGCTGCTGCTG 349
DB 427 TCTCTCAGCTGTGTACATGCTGAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 486
QY 350 CGAACGCGC-----AGAAATGCTACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 403
DB 487 CCAGCCCGCAGTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 546
QY 404 AGAGATCTGAGTACGCTGTATGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 463
DB 547 ACCAGAGTGTGAGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 606
QY 464 GAGGGAAGACAGAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 523
DB 607 TCGAGAGAGGGGAGAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 666
QY 524 GGTACTTGCAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 583
DB 667 AGCTCTTCAAGGATTAATCTCTGCGGCGCAGATCCGCTGCTGCTGCTGCTGCTG 726
QY 584 GTCTCTACACCACTCTGCAAAATTCAGTAGTGTAGAGAAACCGCTGAGNCCAGTT 643
DB 727 GTCTCTACAGAAAGTCTGCAAAATGATGTGATGATGATGATGATGATGATGAT 786
QY 644 A 644
DB 787 A 787

RESULT 12
US-08-684-862-11
Sequence 11, Application US/08684862
Patent No. 5759541
GENERAL INFORMATION:
APPLICANT: Bach, Alfred
APPLICANT: Hillebrand, Heinz
APPLICANT: Hillebrand, Siegfried
TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Keil & Weinkeuf
STREET: 1101 Connecticut Avenue
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5 25 inch, 360 Kb storage
COMPUTER: IBM AT-compatible, 80286 processor
OPERATING SYSTEM: MS-DOS version 5.0
SOFTWARE: WordPerfect version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/684,862
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/361,705

FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/966,040
FILING DATE: 30-DEC-1992
APPLICATION NUMBER: PCT/EP91/01361
FILING DATE: 19-JUL-1991
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 957 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Agkistrodon rhodostoma
FEATURE:
LOCATION: 210 to 911
OTHER INFORMATION: the coding region shown in (2)(1x)(B)
OTHER INFORMATION: codes for the protein of SEQ ID NO: 6
US-08-684-862-11

Query Match 8.2%; Score 95.8; DB 3; Length 957;
Best Local Similarity 49.9%; Pred. No. 5,3e-17;
Matches 297; Conservative 0; Mismatches 289; Indels 9; Gaps 2;

QY 40 TTGTTCTGCTCGGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 99
DB 279 TTTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 338
QY 100 CAGAACTCTTACACATTCGCGGCTGCGCTGCGACAGTCTTGAAGCCGACCAAGAGCCAGG 159
DB 339 AGTACGATCTCAAGATGAGATTTGATGATGATGATGATGATGATGATGATGATGATG 398
QY 160 AGCAGATGTGAGGCGCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 219
DB 399 CAGACAGAGAAAGCAAGAGAAAGTCAATTTGCTCCATTAAGAAACGATGAGTACTG 458
QY 220 GCTAAGAGCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 279
DB 459 GACAGAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 518
QY 280 AGCATGAGATTTGCTGCGAGTGCCTTACCGCGGAGGAGACTTGTGCTGCTGCTGCTG 339
DB 519 CCTCTGAGCTTGCCTTCCAGCGCTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 578
QY 340 GGTCTGCTGCGAGAGGCGAATGCTTACC-----GTCTGCTGCTGCTGCTGCTGCTG 393
DB 579 GGCTCAATCACACCTGTTAAAGTGAATTTCCCGATGCTCCCTCATTTGCTTAACATTAC 638
QY 394 GTGCTGTGAGAGANGTCTGAGTAGCTTA--TGACCCGCTGACACACCCGAGCAG 450
DB 639 CTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 698
QY 451 TTCTGCGCGCGGAGAGAGCAAGACAGAGACTCTGCAAGAGTGTGCTGCTGCTGCTGCTG 510
DB 699 TTCTGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 758
QY 511 CTGATCTGCAAGCGGTACTGCAAGGCGCTTGTCTTTGGAGAAAGCCCGTGTGGCCAA 570
DB 759 CTGATCTGTAATGACAAAGTCCAGGCGCATTTATCTTATGAGGCGGCGCATCTTGTG 818
QY 571 CTGCGGCTGCGAGTGTCTACACCAACCTCTGCAATTTACGAGAGGATAGAGA 625
DB 819 CCTCTTAAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 873

RESULT 13
US-08-358-782D-14
Sequence 14, Application US/08358782D
Patent No. 5674682
GENERAL INFORMATION:
APPLICANT: Croce, Carlo

APPLICANT: Gomella, Leonard
APPLICANT: Mulholland, S. Grant
APPLICANT: Moreno, Jose
APPLICANT: Fischer, Rainer
TITLE OF INVENTION: Methods of Detecting Micrometastasis of Prostate
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
STREET: One Liberty Place 46th. Floor
CITY: Philadelphia
STATE: PA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,782D
FILING DATE: 15-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Beardell, Lori Y.
REGISTRATION NUMBER: 34,293
REFERENCE/DOCKET NUMBER: TJU-1327
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1462 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-358-782D-14

Query Match 7.48; Score 85.8; DB 2; Length 1462;
Best Local Similarity 49.68; Pred. No. 3e-14;
Matches 336; Conservative 0; Mismatches 300; Indels 42; Gaps 3;

2 CGGAGCCCTGGAGGGGCGGCACTGTCATGGAAGCAATTTCTGCTGGGCTCTCTGG 61
128 CCCAACCTGGAGGCTCTGTGCTCTCTGTCGAGGGGAGCTGTGGCGGTCTCTGG 187
62 TGCATCCGAGTGGGCTGTCAGCCGACACAGCTTTTCAGAACTCTACACATCGGGC 121
188 TGCACCCCGAGTGGGCTCTCTGACAGCTGCCCATCATCAGAAACCAAGCTGATCTTGC 247
122 TGGGCTGACAGCTTTGAGGCGGACCAAGAGCCAGGAGGAGGAGGAGGAGGAGGAGG 181
248 TGGGCTGACAGCTTTGAGGCGGACCAAGAGCTTTGAGGAGGAGGAGGAGGAGGAGG 304
182 TCTCCGTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 224
305 ACAGCTTCCACACACCCCTCTAGCATATGAGCCCTCTGGAAGATTCATCTCTAGGGCAG 364
224 -----ACGACCTCATGCTCATCAAGTTGAGAGCAATCCGTGCGAGTCTG 268
365 GTGATGACTCCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 424
269 ACACCATCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 328
425 ATGCTGTGAAGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 484
329 TTTGCTGCTGGG-----TCTGTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 382
485 CCTAGGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 544
383 TGAAGTGTGCTGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 442

DB 545 TGGACCTCCATGTTATTTCCAAAGAGCTGTGTCGCAAGTTCACCTTCAGAAAGTGACCA 604
443 CAGCATGCTCTGCGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 502
605 AGTTCATGCTGTGTGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 664
503 GGGGCCCCCTGATCTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 562
665 GGGGCCCCCTGATCTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 724
563 GTGGCCCACTGGCGGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 622
725 GTGGCCCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 784
623 AGAAACCGTCCAGNCCA 640
785 AGGACACCATCGTGCCCA 802

RESULT 14

US-08-358-782D-13
Sequence 13, Application US/08358782D
Patent No. 5674682

GENERAL INFORMATION:

ATTORNEY/AGENT INFORMATION:

APPLICANT: Gomella, Leonard
APPLICANT: Mulholland, S. Grant

APPLICANT: Moreno, Jose

TITLE OF INVENTION: Methods of Detecting Micrometastasis of Prostate

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &

STREET: One Liberty Place 46th. Floor
CITY: Philadelphia

STATE: PA
ZIP: 19103

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/358,782D
FILING DATE: 15-DEC-1994

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Beardell, Lori Y.
REGISTRATION NUMBER: 34,293

REFERENCE/DOCKET NUMBER: TJU-1327
TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:

LENGTH: 992 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO

ANTI-SENSE: NO
US-08-358-782D-13

Query Match 7.28; Score 84.2; DB 2; Length 992;
Best Local Similarity 49.48; Pred. No. 6.9e-14;
Matches 335; Conservative 0; Mismatches 301; Indels 42; Gaps 3;

2 CGGAGCCCTGGAGGGGCGGCACTGTCATGGAAGCAATTTCTGCTGGGCTCTCTGG 61
147 CCCAACCTGGAGGCTCTGTGCTCTCTGTCGAGGGGAGCTGTGGCGGTCTCTGG 206

Wed Sep 29 14:27:29 1999

us-09-030-606-175.rni

Page 11

Db 1066 AGGACACCATCGTGCCA 1083

Search completed: September 28, 1999, 11:34:10
Job time: 2130 sec

This Page Blank (uspto)

OM of: US-09-030-606-175 to: PIR-60.* out_format : pfs
Date: Sep 25, 1999 6:28 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:

-MODEL=frame+2p.model -DEV=xlp
-O=/cgn2.1/USPTC.spool/US09030606/runat.24091999.171617.29825/app-query.fasta.1
-DB=PIR-60 -QFM=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.500
-MINMATCH=0.100 -LOPCL=0.000 -LOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=bloms62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=score -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=stat -USER=US09030606 -NCPU=6 -ICPU=3 -WAIT
-THREADS=1

Search information block:

Query: US-09-030-606-175
Query length: 1167
Database: PIR-60.*
Database sequences: 122810
Database length: 40065486
Search time (sec): 182.540000

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
PIR2:A53968	+	530.50	6.1e-24	253	serine proteinase SCCE precursor
PIR2:I56559	+	498.50	5.27.05	260	neuropsin - mouse
PIR2:A44284	+	478.00	5.06.68	244	tissue kallikrein (EC 3.4.21.35)
PIR1:NM5065	+	472.00	4.99.96	261	7S nerve growth factor gamma ch
PIR2:A27547	+	470.50	4.98.76	248	trypsin (EC 3.4.21.4) I precursor
PIR2:A25067	+	469.50	4.97.75	247	trypsin (EC 3.4.21.4) precursor
PIR2:S55066	+	468.50	4.96.68	248	trypsin (EC 3.4.21.4) I precursor
PIR2:S45303	+	467.50	4.95.64	248	trypsin (EC 3.4.21.4) II precursor
PIR2:A34079	+	466.00	4.93.71	261	tissue kallikrein (EC 3.4.21.35)
PIR2:S31136	+	463.00	4.90.59	261	tissue kallikrein (EC 3.4.21.35)
PIR2:EB0236	+	463.00	4.90.65	259	tissue kallikrein (EC 3.4.21.35)
PIR2:Q01472	+	462.00	4.89.55	261	tissue kallikrein (EC 3.4.21.35)
PIR2:S33772	+	462.00	4.89.98	246	trypsin (EC 3.4.21.4) V precursor
PIR2:A31136	+	461.50	4.89.14	257	tissue kallikrein (EC 3.4.21.35)
PIR2:S01971	+	461.00	4.88.51	261	tissue kallikrein (EC 3.4.21.35)
PIR1:TRDGC	+	461.00	4.88.51	261	tissue kallikrein (EC 3.4.21.35)
PIR2:Q01471	+	459.00	4.87.38	246	trypsin (EC 3.4.21.4) precursor
PIR2:A41020	+	458.00	4.86.86	246	trypsin (EC 3.4.21.4) V precursor
PIR2:A29746	+	458.00	4.85.39	259	tissue kallikrein (EC 3.4.21.35)
PIR2:S05494	+	457.00	4.85.45	247	tissue kallikrein (EC 3.4.21.35)
PIR1:TRB07R	+	456.50	4.84.75	229	trypsin (EC 3.4.21.4) IV precursor
PIR2:A29745	+	456.00	4.84.77	249	trypsin (EC 3.4.21.4) precursor
PIR2:A29745	+	455.00	4.79.15	261	tissue kallikrein (EC 3.4.21.35)
PIR1:EGMSB	+	450.00	4.77.07	261	tissue kallikrein (EC 3.4.21.35)
PIR2:A29586	+	449.00	4.76.03	261	tissue kallikrein (EC 3.4.21.35)
PIR2:A32297	+	449.00	4.76.03	261	prostate-specific serine protease
PIR2:S25528	+	448.50	4.75.93	246	trypsin (EC 3.4.21.4) precursor
PIR1:KORITN	+	448.00	4.75.04	259	trypsin (EC 3.4.21.4) precursor
PIR1:KORITN	+	447.00	4.73.84	265	tissue kallikrein (EC 3.4.21.35)
PIR2:A25606	+	446.00	4.72.91	261	tissue kallikrein (EC 3.4.21.35)
PIR2:A35871	+	445.00	4.72.38	243	trypsin (EC 3.4.21.4) precursor
PIR2:T01779	+	443.50	4.70.61	250	trypsin (EC 3.4.21.4) precursor
PIR1:TRRRT1	+	442.50	4.69.69	246	trypsin (EC 3.4.21.4) I precursor
PIR2:S31779	+	442.50	4.69.93	238	trypsin (EC 3.4.21.4) III precursor
PIR1:TRMS43	+	442.00	4.68.74	261	tissue kallikrein (EC 3.4.21.35)
PIR2:TS3813	+	441.50	4.68.74	247	trypsin (EC 3.4.21.4) precursor
PIR1:TRPGTR	+	440.50	4.68.06	231	trypsin (EC 3.4.21.4) precursor
PIR1:TRDGS	+	438.50	4.65.52	229	trypsin (EC 3.4.21.4) precursor
PIR1:KQMS1	+	437.00	4.65.52	261	tissue kallikrein (EC 3.4.21.35)
PIR1:KQMS1	+	436.50	4.63.99	262	tissue kallikrein (EC 3.4.21.35)
PIR2:S35711	+	435.00	4.61.46	261	tissue kallikrein (EC 3.4.21.35)

PIR2:A25852 + 434.50 461.34 2.2e-18 247 I trypsin (EC 3.4.21.4) I prec
PIR2:S31778 + 434.00 461.30 2.4e-18 231 I trypsin (EC 3.4.21.4) II pre
PIR2:S49489 + 430.00 456.80 4.1e-18 242 I trypsin (EC 3.4.21.4) precu
seq.name: PIR2:A53968

seq.documentation_block:

serine proteinase SCCE precursor - human
N:Alternate names: stratum corneum chymotryptic enzyme
C:Species: Homo sapiens (man)
C:Date: 07-Jul-1995 #sequence-revision 07-Jul-1995 #text-change 24-Sep-1998
C:Accession: A53968
R:Hansson, L.; Stromqvist, M.; Baeckman, A.; Wallbrandt, P.; Carlstein, A.; Egellrud, J. Biol. Chem. 269, 19420-19426, 1994
A:Title: Cloning, expression, and characterization of stratum corneum chymotryptic en
A:Reference number: A53968
A:Accession: A53968
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-253 <HAN>
A:Cross-references: GB:L33404; NID:9521214; PID:9532504
C:Genetics:
A:Gene: GDB:PRSS6; SCCE
A:Cross-references: GDB:377730
A:Map position: 7q35-7q35
C:Superfamily: trypsin; trypsin homology
F:30-245/Domain: trypsin homology <TRI>

alignment_scores:

Quality: 530.50 Length: 214
Ratio: 3.255 Gaps: 3
Percent Similarity: 76.168 Percent Identity: 47.196

alignment_block:

US-09-030-606-175 x A53968 ..
Align seg 1/1 to: A53968 from: 1 to: 253

```

1 GCGCAGCCCTGGCAGCGGCGCTGTCATGMAAAGATTGTCGTC 50
   ::::::::::::::::::::  ::::::::::::::::::::
40 SerisProtrpGlnValAlaLeuLeuSerGlyAsnInLeuHsCysG1 56
51 GGGCGTCCTGGTCGCGGAGGCGGCGTCGACCGGACAGCTTCC 100
   ::::::::::::::::::::  ::::::::::::::::::::
56 YGlyValLeuValAsnGluArgTyrPylAlaAlaHsCysGlysm 73
101 AGAAGCTCTACACCATCGGCGCTGGCGCTGCACAGTCTTGAGCGACAA 150
   ::::::::::::::::::::  ::::::::::::::::::::
73 etAnsgluTyrThrValHsLeuGlySerSphrLeu...GlyAspArg 88
151 GAGCCAGGAGCCAGATGTGGAGGCGGCGCTTCCGTCGACGCCAGCA 200
   ::::::::::::::::::::  ::::::::::::::::::::
89 Arg.....AlaGlnArgIleLysAlaSerLysSerPheArgHsProG1 103
201 GTAAAGAGACTGTGCTGCTAAGCAAGCTCATGCTATGAGTGGACG 250
   ::::::::::::::::::::  ::::::::::::::::::::
103 YTYSerThrGlnThrHsValAsnAspLeuMetLeuValLysLeuAsn 120
251 AATCGGTGTCGAGTGCACACCATCGGAGCATACATGCTGTCGACG 300
   ::::::::::::::::::::  ::::::::::::::::::::
120 erGlnAlaArgLeuSerSerMetValLysValArgLeuProSerArg 156
301 TGCCCTACCGCGGGAACCTTGCTGCTGCTGCTGCTGCTGCTGCTG 350
   ::::::::::::::::::::  ::::::::::::::::::::
137 CysgluProProGlyThrThrCysThrValSerGlyTyrPylThrThr 153
351 GAAGCGG.....AGATGCTACCGCGTCGTCGCTGCTGCTGCTGCTG 394
153 rSerProAspValThrPheProSerAspLeuMetCysValAlaPylLys 170
395 TGGTGTCTGAGGANGTCTGCAGTACGCTCATGACCGGCTGTACACCC 444
   ::::::::::::::::::::  ::::::::::::::::::::

```

```

170 euileserProglinaSPcysThrlyValTyrlysaSpleuleuGluasn 186
445 AGCATGTTCTGCGCGCGGAGGCAAGACAGACAGTCTGCAACGG 494
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
187 SerMetleucylsAlaGlyIleProaspSerlysaSnaAlacysasnG1 203
495 TGACTGTGGGGGCGCCCTGATCTGCAACGGGCTACTTGACAGGCGCTGTGT 544
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203 yaSPserGlyGlyProleuValCysAlaGlyThrleuGlnGlyleuValS 220
545 CTTTCGGAAAAAGCCCGTGTGGCCACTTGCGCTGCCAGGTGTCTACAC 594
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 extirpGlyThrPheProCysGlyGlnProAsnAspProGlyValTyrThr 236
595 AACCTCGCAATTCACATGATGATGATAGATAAGAAAACGCTCCAG 636
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
237 GlnValCysLysPheThrlySTripIleAsnAspThrMetLys 250
seq_name: p1r2:156559

```

```

seq_documentation_block:
C:Species: Mus musculus (house mouse)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 08-Sep-1997
C:Accession: 156559
R:Chen, Z.L.; Yoshida, S.; Kato, K.; Momota, Y.; Suzuki, J.; Tanaka, T.; Ito, J.; Nishit
J. Neurosci. 15, 5088-5097, 1995
A:Title: Expression and activity-dependent changes of a novel limbic-serine protease gen
A:Reference number: 156559; MUID:95348617
A:Accession: 156559
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-260 <RES>
A:Cross-references: GB:D30785; NID:g1020090; PID:g1020091
C:Superfamily: trypsin; trypsin homology
F:33-252/Domain: trypsin homology <TRI>

```

```

alignment_scores:
Quality: 498.50 Length: 216
Ratio: 2.950 Gaps: 4
Percent Similarity: 78.241 Percent Identity: 41.667

```

```

alignment_block:
US-09-030-606-175 x 156559 ..

```

```

Align seg 1/1 to: 156559 from: 1 to: 260

```

```

1 GGCAGAGCCCTGGCAGCGGCGACTGTCATGAGAAAACGAATGTCTGCTC 50
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
43 SerGlnProtrpIleAlaIleAlaLeuPheGlnGlyGlnArgleuIleCysG1 59
51 GGGCGTCTCTGTCATCCGAGTGGGTGCTGTCAGCCGACACTGTTCC 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59 yGlyValIleValGlyAspArgTyrValIleThrAlaAlaHisCysLysL 76
101 AGAACCTCTACCATCGGCGCTGCACAGTCTTGAGCGGACCA 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 yseGlnLysTyrSerValArgleuGlyAspHisSerleuGlnSerHisArgasp 92
151 GAGCCAGAGGACAGATGTGAGAGCGAGCTCTCCGTACGACCCAGA 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
93 GlnPro...GlnGlnGlnIleGlnValAlaGlnSerIleGlnHisProcy 108
201 GTCACAGACTCTTG.....CTCGCTACGACCTATGCTATCA 241
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
108 sTyrAsnAspSerAsnProGlnAspHisSerHisAspIleMetleuIleA 125
242 AGTTGAGCATCCGATCGATCGAGTGTGACACATCCGAGACATG 291
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 rglGlnGlnAsnSerAlaAsnleuGlyAspLysValLysProValGlnleu 141
292 GCCTGCGAGTGCCTTACCGCGGGGAACTCTTGCTCTGCTGCTGGG 341

```

```

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
142 AlaAsnleuCysProLysValGlyGlnLysCysIleIleSerGlyTrp1 158
342 TCTGTGGGGAAGGCGAGA.....ATGCTTACCGTGTGCTGCACTGCGTCA 385
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
158 yThrValThrSerProGlnGlnAsnPheProAsnThrleuAsnCysAlaG 175
386 ACGTGCGGTGGTGTGTGAGANGTGTGAGTGAAGTCAATACCCGGT 435
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
175 luvAllyIleTyrSerGlnAsnLysCysGlnArgAlaTyrProGlyLys 191
436 TACACCCAGCATGTCTTGCCGCGCGGAGGACAGACAGACAGACTC 485
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
192 IleThrGlnGlyMetValCysAla...GlySerSerAsnGlyAlaAspTh 207
486 CTGCACAGGTGACTCTGGGGGCGCCCTGATCTGCAACGGGCTGTGCAAG 535
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 rcysGlnGlyAspSerGlyGlyProleuValCysAspGlyMetleuGlnG 224
536 GCCTGTGCTCTTTCGGAAGCCGCTGTGGCCACTTGCGCTGCCAGGT 585
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
224 lylleThrSertrpGlySerAspProCysGlyLysProGlnLysProGly 240
586 GTCTACACCAACCTCTGCAATTCATGATGATAGATAAGAAAACGCTC 633
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 ValTyrThrlyleCysArgTyrThrThrTrpIleLysLysThrMet 256
seq_name: p1r2:A44284

```

```

seq_documentation_block:
tissue kallikrein (EC 3.4.21.35) homolog - rat (fragment)
N:Alternate names: proteinase B light chain
C:Species: Rattus norvegicus (Norway rat)
C:Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 29-Jan-1999
C:Accession: A44284; C41429
R:Ma, J.X.; Chao, J.; Chao, L.
Biochemistry 31, 10922-10928, 1992
A:Title: Molecular cloning and characterization of RKLK10, a cDNA encoding T-kininoge
A:Reference number: A44284; MUID:93041794
A:Accession: A44284
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-244 <MAL>
A:Cross-references: GB:S48142; NID:g259430; PID:g259431
A:Experimental source: submandibular gland
A>Note: sequence extracted from NCBI database (NCBI:118095, NCBI:118096)
R:Kato, H.; Nakamishi, E.; Enjoji, K.; Hayashi, I.; Oh-Ishi, S.; Iwanaga, S.
J. Biochem. 102, 1389-1404, 1987
A:Title: Characterization of serine proteinases isolated from rat submaxillary gland:
A:Reference number: A41429; MUID:88198057
A:Accession: C41429

```

```

A:Status: preliminary
A:Molecule type: protein
A:Residues: 10-61 <KAT>
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:10-236/Domain: trypsin homology <TRY>
F:48,103,196/Active site: His, Asp, Ser #status predicted

```

```

alignment_scores:
Quality: 478.00 Length: 225
Ratio: 2.987 Gaps: 4
Percent Similarity: 71.111 Percent Identity: 41.333

```

```

alignment_block:
US-09-030-606-175 x A44284 ..

```

```

Align seg 1/1 to: A44284 from: 1 to: 244

```

```

1 GGCAGAGCCCTGGCAGCGGCGACTGTCATGAGAAAACGAATGTCTGCTC 50
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
20 SerGlnProtrpIleValAlaIleIle.....AsnGlyLysleuGlnCysG1 34

```

```

51 GGGCGCTGCTGTCATCCGACAGTGGCTGTGCACGCGACACTGTTCC 100
|||||..... |||..... |||..... |||..... |||.....
34 yglYValLeuLeuAspProSerTrpValLeuThrAlaAlaHisCysTyr 51
|||||..... |||..... |||..... |||..... |||.....
101 AGAATCTCTACACCATCGGCTGGGCTGGCTGACAGTGTAGCGGACCA 150
|||||..... |||..... |||..... |||..... |||.....
51 exAsnTyrTrpHisValLeuLeuGluArgAsnAsnLeuPheGluAsp... 66
|||||..... |||..... |||..... |||..... |||.....
151 GAGCGGAGGAGCAGATGGTGGAGCGGACGCTCTCCGTACGGACCCAGA 200
|||||..... |||..... |||..... |||..... |||.....
67 GluProPheAlaGlnIleValArgPheValAsnGlnSerPheTrpHisPro 83
|||||..... |||..... |||..... |||..... |||.....
201 GTACACAGACTTGTCTC..... 219
|||||..... |||..... |||..... |||..... |||.....
83 pYrTrpSerProPheLeuMetAlaArgAsnHisThrArgGlnArgGlyAspAsp 100
|||||..... |||..... |||..... |||..... |||.....
220 ..GCTAACGACCTCATGCTCATCACTGACAGCAATCGCTGCCAGTCT 267
|||||..... |||..... |||..... |||..... |||.....
100 ySerAsnAspLeuMetLeuLeuHisLeuSerGluProAlaAspIleThr 116
|||||..... |||..... |||..... |||..... |||.....
268 GACACCATCCGAGACATGACATGCTGTGGACATGCTCCATCCGCGGGA 317
|||||..... |||..... |||..... |||..... |||.....
117 AspGlyValLeuValLeuAspLeuProThrGluGluProLysValGlySe 133
|||||..... |||..... |||..... |||..... |||.....
318 CTCTTGCCTCTGTTCTGCTGGGCTGCTGCTG.....GCGAACGCGCAGA 361
|||||..... |||..... |||..... |||..... |||.....
133 ThrCysLeuAlaSerGlyTrpProLysThrCysProLeuAsnTrpGlu 150
|||||..... |||..... |||..... |||..... |||.....
362 TCCCTACCGCTGCTGACATCGCTGACAGTGTGGTGTGTGACAGAGTGC 411
|||||..... |||..... |||..... |||..... |||.....
150 euProAspAspLeuGlnCysValAsnIleHisLeuSerAsnGluLys 166
|||||..... |||..... |||..... |||..... |||.....
412 TGCAGTAACCTATGACACCGCGCTGTACACCCAGCATTTCTGCCCGG 461
|||||..... |||..... |||..... |||..... |||.....
167 CysIleGluAlaTyrLeuIleGlnLysValThrAspLeuMetLeuCysAla 183
|||||..... |||..... |||..... |||..... |||.....
462 CGAGAGGCGACAGACAGACGACTCTGACAGGCTGCTGGGGGCGCC 511
|||||..... |||..... |||..... |||..... |||.....
183 yGluMetAspGlyArgLysAspThrCysLysGlyAspSerGlyGlyPro 200
|||||..... |||..... |||..... |||..... |||.....
512 TGATGTGACAGCGGTACTTGACAGGCGCTGTGTCTTTCGAAAACCCG 561
|||||..... |||..... |||..... |||..... |||.....
200 euIleCysAspGlyValLeuGlnGlyIleThrSerTrpGlyAsnValPro 216
|||||..... |||..... |||..... |||..... |||.....
562 TGTGGCCAACTTGGCGTGGCAGAGTGTCTACACCACTCTGCAAAATTC 611
|||||..... |||..... |||..... |||..... |||.....
217 CysAlaGluProTyrAsnProGlyValTyrThrLysLeuIleLysPheThr 233
|||||..... |||..... |||..... |||..... |||.....
612 TGAGTGATAGAGAAACCGTCCAG 636
|||||..... |||..... |||..... |||..... |||.....
233 rSerTrpIleLysGluValMetLys 241
|||||..... |||..... |||..... |||..... |||.....
seq_name: p1r1:NGMSG
seq_documentation_block:
7S nerve growth factor gamma chain (EC 3.4.21.-) precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 18-Dec-1981 #sequence_revision 17-May-1985 #text_change 20-Mar-1998
C:Accession: A91005; A90949; A93510; A92341; A00942; A21093; A22705
R:Evans, B.A.; Richards, R.I.
EMBO J. 4, 133-138, 1985
A:Title: Genes for the alpha and gamma subunits of mouse nerve growth factor are contig
A:Reference number: A91005; MUID:85257431
A:Accession: A91005
A:Molecule type: DNA
A:Residues: 1-261 <EVA>
R:Ullrich, A.; Gray, A.; Wood, W.I.; Hayflick, J.; Seeburg, P.H.
DNA 3, 387-392, 1984
A:Title: Isolation of a cDNA clone coding for the gamma-subunit of mouse nerve growth fa
A:Reference number: A90949; MUID:85076169
A:Accession: A90949
A:Molecule type: mRNA

```

```

A:Residues: 1-261 <DUL>
A:Cross-references: GB:X01389; NID:953373; PID:953374
R:Howles, P.N.; Dickinson, D.P.; DiCiprio, L.L.; Woodworth-Gutai, M.; Gross, K.W.
Nucleic Acids Res. 12, 2791-2805, 1984
A:Title: Use of a cDNA recombinant for the gamma-subunit of mouse nerve growth factor
A:Reference number: A93510; MUID:84169573
A:Accession: A93510
A:Molecule type: mRNA
A:Residues: 127-202, 'E', 204-261 <HOW>
A:Cross-references: GB:X00472; NID:954260; PID:954261
A:Experimental source: Inbred strain DBA/2J
R:Thomas, K.A.; Baglan, N.C.; Bradshaw, R.A.
J. Biol. Chem. 256, 9156-9166, 1981
A:Title: The amino acid sequence of the gamma-subunit of mouse submaxillary gland 7 S
A:Reference number: A92341; MUID:81264363
A:Accession: A92341
A:Molecule type: protein
A:Residues: 25-107, 112-261 <THO>
A:Experimental source: Outbred strain Swiss Webster
C:Comment: 7S nerve growth factor is composed of two alpha chains, a beta dimer compo
C:Comment: The active form of the gamma chain occurs naturally as combinations of elc
C:Genetics:
A:Map position: 7
A:Introns: 16/1; 69/2; 165/1; 210/3
C:Superfamily: trypsin; trypsin homology
C:Keywords: glycoprotein; growth factor; hydrolase; serine proteinase; submandibular
F:1-18/Domain: signal sequence #status predicted <SIG>
F:25-253/Domain: signal homology <TRX>
F:25-107/Domain: trypsin homology <B1>
F:25-107/Domain: segment B1 <B1>
F:25-107, 112-261/Product: nerve growth factor gamma chain (active form) #status exper
F:112-261/Domain: segment A <GAX>
F:112-164/Domain: segment C <GCC>
F:165-261/Domain: segment B2 <GCC>
F:31-173, 50-66, 152-219, 184-198, 209-234/Disulfide bonds: #status predicted
F:65, 120, 213/Active site: His, Asp, Ser #status predicted
F:102/Binding site: carbohydrate (Asn) (covalent) #status experimental

alignment_scores:
Quality: 472.00 Length: 228
Ratio: 2.969 Gaps: 4
Percent Similarity: 69.737 Percent Identity: 39.912

alignment_block:
US-09-030-606-175 x NGMSG ..
Align seg 1/1 to: NGMSG from: 1 to: 261
1 GCGCAGCCCTGGCAGCGGCGACCTGTCATGGAAGAAAGATGTTCTGCTC 50
|||||..... |||..... |||..... |||..... |||.....
35 SerGlnProTrpHisValAlaValTyrArgTyrThrGlnIleTyrLeuGly 51
|||||..... |||..... |||..... |||..... |||.....
51 GGGCGTCTGTGTCATCCGACAGTGGGCTGTGTCACGCGCACACTGTTTC 100
|||||..... |||..... |||..... |||..... |||.....
51 yGlyValLeuLeuAspProAsnTrpValLeuThrAlaAlaHisCysTyr 68
|||||..... |||..... |||..... |||..... |||.....
101 AGAATCTCTACACCATCGGCTGGGCTGCGCACAGCTTGAGGCGGACCA 150
|||||..... |||..... |||..... |||..... |||.....
68 spAspAsnTyrLysValTyrLeuGlyLysAsnAsnLeuPheLysAsp... 83
|||||..... |||..... |||..... |||..... |||.....
151 GAGCGGAGGAGCAGATGGTGGAGGCGGCGCTTCGCCAGGACCCAGA 200
|||||..... |||..... |||..... |||..... |||.....
84 GluProSerAlaGlnHisArgPheValSerLysAlaIleProHisProG 100
|||||..... |||..... |||..... |||..... |||.....
201 GTACACAGACTTGTCTC..... 219
|||||..... |||..... |||..... |||..... |||.....
100 yPheAsnMetSerLeuMetArgLysHisIleArgPheLeuGluIuTyrAsp 117
|||||..... |||..... |||..... |||..... |||.....
220 ..GCTAACGACCTCATGCTCATCAAGTTGAGACGAATCCGTGCGAGTCT 267
|||||..... |||..... |||..... |||..... |||.....
117 ySerAsnAspLeuMetLeuLeuArgLeuSerLysProAlaAspIleThr 133
|||||..... |||..... |||..... |||..... |||.....
268 GACACCATCCGAGACATGACATGCTTGCAGAGTCCCTACCGGCGGGA 317
|||||..... |||..... |||..... |||..... |||.....

```


alignment_scores:
 Quality: 469.50 Length: 211
 Ratio: 2.953 Gaps: 4
 Percent Similarity: 75.355 Percent Identity: 41.706

alignment_block:
 US-09-030-606-175 x A27547 ..

Align seg 1/1 to: A27547 from: 1 to: 247

```

7 CCTGTGGCAGCGGCACATGCTCATGGAAGCAATGTTCTGCTGGCGCT 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
37 PRTYRGLNValSerLeu...AsnAlaGlyTyrHisPheCysGlyLeu 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
57 CCTGTGACATCCGAGTGGTGTGCTGACACCCGACACATGTTCCAGAACT 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
52 rleuLeuSerGlnTyrValSerLeuAlaHisCysTyrLysSer 69
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
107 CCTACACCATCGGCGCTGGCGCTGCACAGTCTTGAGCCGACCAAGACCA 156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
69 rglleGlnValArgLeuGlyGlnHisAsnIleAspVal...ValGluGly 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
137 GGGACCCAGATGCTGGAGCCAGCCCTCCGTAACGACCCAGAGTACA 206
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85 GlyGluGlnPheIleAspAlaIleAlaLysIleIleArgHisProSerLys 101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 CAGACTCTGCTGGCTACGACCTCATGCTCATGAGTGGAGCAATCCG 256
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 nAlaAsnTyrPheAspAsnAspIleMetLeuIleLysLeuAsnSerPro 118
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
257 TGTCCGAGTGTGACACCATCCGAGCATCAGATGCTTCCGAGTCCCT 306
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
118 lAtThrLeuAsnSerArgValSerThrValSerLeuProArgSerGly 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
307 ACCCGGGGAACTCTTGCTGCTGTTGCTGGCGT.....CTGCTGGC 350
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
135 SerSerGlyThrLysCysLeuValSerGlyTyrGlyAsnThrLeuSer 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
351 GAACGGGCAATGCTACCGTGTGACGCTGACGCGTAACGCTGGGCTGT 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 rGlyThrAsnTyrProSerLeuLeuGlnCysLeuAspAlaProValLeu 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
401 CTGAGGANGTCTGACATAAGTCTATGACCCGCTGACCAACCCACATG 450
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
168 eAspSerSerCysLysSerSerTyrProGlyLysIleThrSerAsnMet 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
451 TTCTGCGCCGCGGAGGCAAGACCAAGACTCTGCGACGCTGACTC 500
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 PheCysLeuGlyPheLeuGlnGlyLysAspSerCysGlnGlyAsp 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
501 TGGGGGGGCGCGATCTGCAAGGGGACTTGAGCGCTTGCTGTG 550
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 rGlyGlyProValValCysAsnGlyGlnSerLeuGlnGlyValValSer 218
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
551 GAAAGCCCGCTGTGGCCAACTTGCGCTGCCAGTGTCTACACCACTC 600
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218 lYTyGly...CysAlaGlnLysGlyLysProGlyValTyrThrLysVal 233
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
601 TGCAAAATTCAGTGAAGTATGAGAAAACCGTC 633
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 CysAsnTyrValAsnTyrPheGlnGlnThrVal 244
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: p1r2:S55067

seq_documentation_block:

trypsin (EC 3.4.21.4) I precursor, pancreatic (clone 1-P38) - chicken
 N:Alternate names: trypsinogen I
 C:Species: Gallus gallus (chicken)
 C:Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 08-Sep-1997
 C:Accession: S55067; S72345; S71155
 R:Wang, K.; Gan, L.; Lee, I.; Hood, L.
 Biochem. J. 307, 471-479, 1995
 A:title: Isolation and characterization of the chicken trypsinogen gene family.

A:Reference number: S55065

A:Accession: S55067

A:Molecule type: mRNA

A:Residues: 1-248 <MAN1>

A:Cross-references: EMBL:U15156

A:Experimental source: clone 1-P38

A:Accession: S72345

A:Molecule type: DNA

A:Residues: 1-248 <MAN2>

A:Cross-references: EMBL:U15156

A:Experimental source: clone 1-P38

R:Wang, K.

submitted to the EMBL data library, September 1994

A:Reference number: S71155

A:Accession: S71155

A:Molecule type: mRNA

A:Residues: 1-102, 'A', 104-248 <MAN>

A:Cross-references: EMBL:U15156; NID:G603904; PID:G603905

A:Experimental source: clone 1-P38

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen

F:1-15/Domain: signal sequence #status predicted <Sig>

F:16-25/Domain: activation peptide #status predicted <AP>

F:26-248/Product: trypsin I #status predicted <Mat>

F:26-241/Domain: trypsin homology <TRY>

F:65,109,202/Active site: His, Asp, Ser #status predicted

alignment_scores:
 Quality: 468.50 Length: 216
 Ratio: 2.910 Gaps: 5
 Percent Similarity: 74.537 Percent Identity: 43.056

alignment_block:
 US-09-030-606-175 x S55067 ..

Align seg 1/1 to: S55067 from: 1 to: 248

```

1 GCGCAGCCCTGGCAGCGGCACATGCTCATGGAAGCAATGTTCTGCTC 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
36 AlaAlaPRTYRGLNValSerLeu...AsnSerGlyTyrHisPheCysG 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 GGGCGCTGCTGTCATCCGAGTGGTGTGCTGACACCCGACATGTTCC 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 yGlySerLeuIleSerSerGlnTyrValSerLeuAlaHisCysTyrL 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 AGAATCTCTACACCATCGGCGCTGCGCTGCACAGTCTTGAGCCGACCA 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
68 ySerSerIleGlnValLysLeuGlyGlyTyrAsnLeuAlaAla...Gln 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 GAGCAGGAGGACCCAGATGCTGGAGCCGCTCTCCGTAACGCGACCCA 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 AspGlySerGlnGlnThrIleSerSerSerLysValIleArgHisSer 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 GTACAACAGACTCTGCTGCTACGACCTCATGCTCATGAGTGGACG 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 YTYrAsnSerAsnThrLeuAsnAsnAspIleMetLeuIleLysLeuSer 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 AATCGGTCCGAGTCTGACACCATCCGAGCATCAGCATGCTCTCGAG 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 yAlaAlaThrLeuAsnSerTyrValAsnThrValProLeuProThrSer 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 TGCCCTACCGCGGGGAACTCTTGCTGCTGTTCTGCTGGGT...CTGCT 347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134 CysValThrAlaGlyThrThrCysLeuIleSerGlyTyrGlyAsnThr 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
348 GGGCAGCGCAGAAATG...CCTACCGTGTGCTGCTGCTGAGACGTCG 394
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150 uSerSerLysLeuTyrProAspValLeuGlnCysLeuAsnAlaProV 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
395 TGTGTCTGAGGANGTCTGACATAAGTCTATGACCCGCTGACACCCC 444
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 AlLeuSerSerSerGlnCysSerSerAlaTyrProGlyArgIleThrSer 183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

445 AGCATGTCGCGCGGCGGAGGCGAAGCAGAGCACTCTCCAGCG 494
    ::::::::::|||:::|||||
184 AsnMetIleCysIleGlyTyrLeuAsnGlyLysAspSerCysGlnG 200
    |||
495 TGACTCTGGGGGCGCCCTGATCTGCACAGGCTACTGCAGGGCCTGTGT 544
    |||
200 yAspSerGlyCysProValValCysAsnGlyGlnLeuGlnGlyPheValS 217
    |||
545 CTTTCGGGAAAGCCCGGTGGCCAGCTTGCGCAGGTGTCTACACC 594
    |||
217 erirpGly...IleGlyCysAlaGlnLysGlyTyrProGlyValTyrThr 232
    ::::::::::
595 AACCTCTGCAGAAATTCAGTACTGATGATAGAGAAACCGTCCAGNCCAGT 642
    ::::::::::
233 LysValCysAsnTyrValSerTyrIleLysThrThrMetSerSerAsn 248
    ::::::::::
seq_name: p1r2:S55066

```

```

seq_documentation_block:
  trypsin (EC 3.4.21.4) II precursor, pancreatic (clone 2-P29) - chicken
N:Alternate names: trypsinogen II
C:Species: Gallus gallus (chicken)
C:Date: 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 08-Sep-1997
C:Accession: S55066; S72347
R:Wang, K.; Gan, L.; Lee, I.; Hood, L.
Biochem. J. 307, 471-479, 1995
A:Title: Isolation and characterization of the chicken trypsinogen gene family.
A:Reference number: S55065
A:Accession: S55066
A:Molecule type: mRNA
A:Residues: 1-248 <MAN1>
A:Cross-references: EMBL:U15157; NID:g603906; PID:g603907
A:Experimental source: clone 2-P29
A:Accession: S72347
A:Molecule type: DNA
A:Residues: 1-248 <MAN2>
A:Cross-references: EMBL:U15157; NID:g603906; PID:g603907
A:Experimental source: clone 2-P29
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-25/Domain: activation peptide #status predicted <APR>
F:26-248/Product: trypsin II #status predicted <MAT>
F:26-241/Domain: trypsin homology <TRY>
F:65,109,202/Active site: His, Asp, Ser #status predicted

```

```

alignment_scores:
  Quality: 467.50      Length: 211
  Ratio: 2.959        Gaps: 5
  Percent Similarity: 74.882      Percent Identity: 42.654

```

```

alignment_block:
  US-09-030-606-175 x S55066 ..

```

```

Align seg 1/1 to: S55066 from: 1 to: 248

```

```

7  CCCGTCGACGGCGGACATGCTCATGGAAGCAATTTCTCTCGGGCGT 56
   |||
38  ProTyrGlnValSerLeu...AsnSerGlyTyrHisPheCysGlyGlySe 53
   |||
57  CCTGTGATCCGACAGTGGTGTGTCTGACGCGACACATGTTCCAGACT 106
   |||
53  rLeuIleAsnSerGlnTyrValLeuSerAlaAlaHisCysTyrTyrSera 70
   |||
107  CCAACACCATCGGGCTGGCGCTGCACAGCTTTGAGCGCGACCAAGCCA 156
   |||
70  rGlieGlnValAlaGlnGlyGlnTyrAsnIleAspVal...GlnGlnLysp 85
   |||
157  GGGAGCCAGATGTGTGAGGCGAGCTCTCCGTACGAGCCAGCCAGATACA 206
   |||
86  SerGlnValValAlaGlySerSerValIleIleAlaGlnHisProLysTyrSe 102

```

```

207  CAGACTCTCTGCTGCAACGACCTCATGCTCATCAAGTTGAGCAATCCG 256
   :
102  rSerIleThrLeuAsnAsnAspIleMetLeuIleLysLeuAlaSerAlaY 119
   |||
257  TGTCCGATGTGACACCATCCGAGACATGACATTCCTTCGAGTGCCTT 306
   |||
119  alGluTyrSerAlaAspIleGlnProIleAlaLeuProSerSerCysAla 135
   |||
307  ACCGCGGGGAAACTTGTCCCTGCTGCTGCTGGGT...CTGTCGCGGAA 353
   |||
136  LysAlaGlyThrLeuLysLeuIleSerGlyTyrGlyPheGlnThrLeuSer 152
   |||
354  CGGC...AGATGCTCTACCTGCTGCACCTGCTGACGCTGCTGCTGTG 400
   |||
152  ngLYTyrAsnTyrProGlnLeuLeuGlnCysLeuAsnAlaProIleLeuS 169
   |||
401  CTGAGANGTCTGCAGTATGACTCTATGACCCGCTGTACACCCAGCATG 450
   |||
169  eraSpGlnGlnCysGlnGlnAlaTyrProGlyAspIleThrSerAsnMet 185
   |||
451  TTTCGCGCGCGGCGGAGGCGAAGCAGAGACTCTGCAACGCTACTC 500
   |||
186  IleCysValGlyPheLeuGlnGlyGlyLysAspSerCysGlnLysPse 202
   |||
501  TGGGGGCGCCCTGATCTGCAACGGGTACTGCGAGGCGCTGTGCTTTG 550
   |||
202  rGlyGlyProValValCysAsnGlyLeuGlnGlnGlyIleValSerTyrG 219
   |||
551  GAAAGCCCGGTGGCCAACTTGGCGTGCAGAGTGTCTACACCAACTTC 600
   |||
219  ly...IleGlyCysAlaLeuLysGlyTyrProGlyValTyrThrLysVal 234
   |||
601  TGCAGATTCAGTACTGATGATAGAGAAACCGTCTC 633
   |||
235  CysAsnTyrValAspTyrIleGlnGlnThrIle 245

```

```

seq_name: p1r2:S45303

```

```

seq_documentation_block:
  tissue kallikrein (EC 3.4.21.35) precursor - dog
C:Species: Canis lupus familiaris (dog)
C:Date: 20-Oct-1994 #sequence_revision 03-Aug-1995 #text_change 17-Mar-1999
C:Accession: S45303; S38487
R:Gauchier, E.R.; Dumas, C.; ChappelaIn, P.; Tremblay, R.R.; Dube, J.Y.
Biochim. Biophys. Acta 1218, 102-104, 1994
A:Title: Characterization of canine pancreas kallikrein cDNA.
A:Reference number: S45303; MUID:94250683
A:Accession: S45303
A:Molecule type: mRNA
A:Residues: 1-261 <GAND>
A:Cross-references: EMBL:X75479; NID:9414018; PID:9414019
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-261/Product: tissue kallikrein #status predicted <MAT>
F:25-253/Domain: trypsin homology <TRY>
F:65,120,213/Active site: His, Asp, Ser #status predicted

```

```

alignment_scores:
  Quality: 466.00      Length: 224
  Ratio: 2.913        Gaps: 3
  Percent Similarity: 71.429      Percent Identity: 39.732

```

```

alignment_block:
  US-09-030-606-175 x S45303 ..

```

```

Align seg 1/1 to: S45303 from: 1 to: 261

```

```

1  GCGAGCCCTGGGACGGCGGACATGCTCATGGAAGCAATTTCTGCTC 50
   ::::::::::
35  SerGlnProTyrGlnAlaAlaLeuTyrHisTyrSerLysPheGlnCysG 51

```

```

51 GGGCGCTGCTGTGATCCGACAGTGGGTCTGTACCCGACATGTTTCC 100
   |||||.....:|||||.....:|||||.....:
51 YGLYVALLLeuValHisProGluTrpValThrAlaHisCysIle 68
101 AGAACTCTACACACATCGGGCTGGGCTGCACAGTCTTGAGGCCACCA 150
   |||||.....:|||||.....:|||||.....:
68 snAspAspIleGluLeuTrpLeuGlyArgTyrAsnLeu...PheGluHis 83
151 GAGCGAGGAGCAGATGTGTGGAGGCGACGCTCCGATACGCGGACCGA 200
   |||||.....:|||||.....:|||||.....:
84 GluSerThrAlaGlnPheValGlnValAlaArgGluSerPheProHisProGln 100
201 GTACAACAGACTCTGTCTGCTGCTAAC.....: 225
   |||||.....:|||||.....:
100 uphAsnLeuSerLeuLeuLysAsnHisThrArgLeuProGluGluAsp 117
226 .....GACCTCATGCTCATCAAGTTGGACGAATCCGCTCCGAGTCT 267
   |||||.....:|||||.....:|||||.....:
117 yrSerHisAspIleMetLeuLeuArgLeuAlaGluProGlnValGlySe 133
268 GACACCATCCGGAGCATCAGCATTCCTCGCAGTCCGACGCGGGGA 317
   |||||.....:|||||.....:|||||.....:
134 AspAlaValArgValLeuAspLeuProThrGlnGluProGlnValGlySe 150
318 CTCTTGCCCTGCTGCTGCTGGGTCTGTCTGCGACGACGAGATG.... 363
   |||||.....:|||||.....:|||||.....:
150 rThrCysTyrAlaSerGlyTrpGlySerIleGluProAspLysPheIle 167
364 ..CCTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 411
   |||||.....:|||||.....:|||||.....:
167 yrProAspAspLeuGlnCysValAspLeuGluLeuLeuSerAsnAspIle 183
412 TGCAGTAGCTCTATGACCCGCTGACACCCGACATGCTTGCGCGG 461
   |||||.....:|||||.....:|||||.....:
184 CysAlaAsnAlaHisSerGlnLysValThrGluPheMetLeuCysAlaGln 200
462 CGAGGCGCAGACGACGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 511
   |||||.....:|||||.....:|||||.....:
200 YHisLeuGluGlyLysAspThrCysValGlyAspSerGlyGlyProL 217
512 TGATCTGCACAGGCTACTTGCAGGCGCTTGTCTTTCGAAAAGCCCG 561
   |||||.....:|||||.....:|||||.....:
217 euIleCysAspGlyValLeuGlnGlyIleThrSerTrpGlyHisValPro 233
562 TGTGGCCCAACTTGGCGTGCAGGCTGTACACCAACGCTGCAATTCAC 611
   |||||.....:|||||.....:|||||.....:
234 CysGlySerProAsnMetProAlaValThrLysValIleSerHisIle 250
612 TGAGTGTATAGAAACCGTC 633
250 uGluTrpIleLeuGluTrpMet 257

```

seq_name: p1r2:A34079

seq_documentation_block:

tissue kallikrein (EC 3.4.21.35) P1 precursor - rat

N:Alternate names: kallikrein-related proteinase k8

C:Species: Rattus norvegicus (Norway rat)

C:Date: 30-Mar-1990 #sequence_revision 30-Mar-1990 #text_change 08-Sep-1997

C:Accession: A34079; S10700

R:Brady, J.M.; Mines, D.R.; MacDonald, R.J.

Biochemistry 28, 5203-5210, 1989

A:Title: Expression of two kallikrein gene family members in the rat prostate.

A:Reference number: A34079; M01D:89352606

A:Accession: A34079

A>Status: Preliminary

A:Molecule type: DNA; mRNA

A:Residues: 1-261

A:Cross-references: GB:M27215; GB:M27216; GB:M27217; NID:g206638; PID:g206640; GB:M27218

A:Experimental source: prostate

R:Elmoujaded, A.; Gutman, N.; Brillard, M.; Gauthier, F.

FEBS Lett. 265, 137-140, 1990

A:Title: Substrate specificity of two kallikrein family gene products isolated from t

A:Reference number: S10698; M01D:90306305

A:Accession: S10700

A:Molecule type: protein

A:Residues: 25-43;112-138 <ELM>

A:Experimental source: submaxillary gland

A>Note: 125-Lys was also found

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; serine proteinase

F:25-253/domain: trypsin homology <TRY>

F:25-111/Product: tissue kallikrein P1 light chain #status experimental <MAT1>

F:112-261/Product: tissue kallikrein P1 heavy chain #status experimental <MAT2>

F:65,120,213/Active site: His, Asp, Ser #status predicted

alignment_scores:

Quality	463.00	Length	225
Ratio	2.987	Gaps	3
Percent Similarity	68.889	Percent Identity	40.889

alignment_block:

US-09-030-606-175 x A34079 ..

Align seg 1/1 to: A34079 from: 1 to: 261

```

1 GCGCAGCCCTGCGACGCGGACATGCTGATGGAACGATGTTCTGCTC 50
   |||||.....:|||||.....:|||||.....:
35 SerGlnProTrpGlnValAlaValThrHisPheAsnLupProGlnCysG 51
51 GGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 100
   |||||.....:|||||.....:|||||.....:
51 YGLYVALLLeuHisProSerTrpValIleThrAlaHisCysTyrS 68
101 AGAACTCTACACATCGGGCTGGGCTGCATGCTGCTGCTGCTGCTGCT 150
   |||||.....:|||||.....:|||||.....:
68 euValAsnTyrGlnValThrPheGlyArgAsnAsnLeuGluAsp... 83
151 GAGCGAGGAGCCAGATGTGTGAGGAGCCGCTGCTGCTGCTGCTGCTGCT 200
   |||||.....:|||||.....:|||||.....:
84 GluProPheAlaGlnHisArgLeuValSerGlnSerPheProHisProGln 100
201 GTACAACAGACTTGTCTC.....: 219
   |||||.....:|||||.....:
100 YPheAsnLeuAspIleLeuLysAsnHisThrArgLysProGlyLysAsp 117
220 ..GCTAACGACTCATGCTCATCAAGTTGGACGACATCCGCTGCTGCT 267
   |||||.....:|||||.....:|||||.....:
117 yrSerAsnAspLeuMetLeuHisIleLysThrProAlaAspIleThr 133
268 GACACATCCGAGCATCAGCATGCTTGCAGAGCCCTACCGCGGGGAA 317
   |||||.....:|||||.....:|||||.....:
134 AspGlyValLysValIleAspLeuProThrGluLupProLysValGlySe 150
318 CTCTTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 361
   |||||.....:|||||.....:|||||.....:
150 rThrCysLeuThrSerGlyTrpGlySerIleThrProLeuLysTrpLup 167
362 TGCCCTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 411
   |||||.....:|||||.....:|||||.....:
167 heProAspAspLeuGlnCysValAsnHisIleLeuSerAspGluLys 183
412 TGCAGTAGCTCTATGACCCGCTGACACCCGACATGCTTGCGCGG 461
   |||||.....:|||||.....:|||||.....:
184 CysIleLysAlaThrAsnAspGluValThrAspValMetLeuCysAlaGln 200
462 CGAGGCGCAGACGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 511
   |||||.....:|||||.....:|||||.....:
200 YGluMetAspGlyLysAspIleCysLysGlyAspSerGlyLysProL 217
512 TGATCTGCACAGGCTACTTGCAGGCGCTTGTCTTTCGAAAAGCCCG 561
   |||||.....:|||||.....:|||||.....:
217 euIleCysAspGlyValLeuGlnGlyIleThrSerTrpGlySerMetPro 233

```



```

201 GTACACAGACTCTTCTC..... 219
    |||||
100 YTrasmetSerLeuMetLeuGlnThrIleProProGlyAlaAsp 117
    |||||
220 ..GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGCCAGTCT 267
    |||||
117 heserAsnAspLeuMetLeuArgLeuSerLysProAlaAspIleThr 133
    |||||
268 GACACCAACCGGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 317
    |||||
134 AspValValLysProIleLeuLeuProThrLysGluProLysProGlyLys 150
    |||||
318 CTCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 361
    |||||
150 rlyscysLeuAlaSerGlyTrpGlySerIleThrProThrArgTrpGln 167
    |||||
362 TGCTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 411
    |||||
167 yspProAspAspLeuGlnCysValPheIleThrLeuLeuProAsnGluAsn 183
    |||||
412 TGCAAGTAGCTCTATGACCCGCTGACACCCACATGTTCTGCGCGG 461
    |||||
184 CysAlaLysValIleTrpLeuGlnLysValIleThrAspValMetLeuCysAlaG1 200
    |||||
462 CGGAGGCGACAGACAGAGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 511
    |||||
200 ygluMetGlyGlyLysAspThrCysArgAspAspSerGlyGlyProL 217
    |||||
512 TGATCTGCACGGGCTACTGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 561
    |||||
217 euileCysAspGlyLysLeuGlnGlyThrThrSerGlyProThrPro 233
    |||||
562 TGTGGCCAACTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 611
    |||||
234 CysGlyLysProGlyValProAlaIleThrThrAsnLeuIleLysPheAs 250
    |||||
612 TGAGTGATAGAGAAACCGTC 633
    |||||
250 nserTrpIleLysAspThrMet 257

```

```

seq_name: p1r2:J01472
seq_documentation_block:
  tryptsin (EC 3.4.21.4) V precursor, b-form - rat
  C:Species: Rattus norvegicus (Norway rat)
  C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 08-Sep-1997
  C:Accession: J01472
  R:Kang, J.; Miesand, U.; Mueller-Hill, B.
  Gene 110, 181-187, 1992
  A:Title: Identification of cDNAs encoding two novel rat pancreatic serine proteases.
  A:Reference number: J01471; MID:9215057
  A:Accession: J01472
  A:Molecule type: mRNA
  A:Residues: 1-246 <RAN>
  A:Cross-references: EMBL:X59013; NID:957414; PID:957415
  A:Experimental source: pancreas
  C:Superfamily: trypsin; trypsin homology
  C:Keywords: hydrolase; protein digestion; serine proteinase; zymogen
  F:1-15/Domain: signal sequence #status predicted <Sig>
  F:16-24/Domain: activation peptide #status predicted <ACT>
  F:25-246/Product: trypsin V, b-form #status predicted <MNT>
  F:25-239/Domain: trypsin homology <TRY>
  F:31-160,49-05,133-233,140-206,171-185/Disulfide bonds: #status predicted
  F:64,108,200/Active site: His, Asp, Ser #status predicted

```

```

alignment_scores:
  Quality: 462.00      Length: 210
  Ratio: 3.000         Gaps: 4
  Percent Similarity: 73.333      Percent Identity: 42.361
alignment_block:

```

```

US-09-030-606-175 x J01472 ..
Align seg 1/1 to: J01472 from: 1 to: 246

7 CCGTGGACGCGGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 56
  |||||
37 ProGlyIleValSerLeuAsnAlaGlySerHisIle...CysGlyLys 52
  |||||
57 CCGTGGACATCCGCGATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 106
  |||||
52 rleuIleThrAspIleTrpValLeuSerAlaAlaHisCysTrpHisProG 69
  |||||
107 CCTACACATCGGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 156
  |||||
69 ILeuGlnValArgLeuGlyGlnHisAsnIle...TrpGluIleGly 84
  |||||
157 GGGAGCCAGATGTTGAGCGCCAGCTTCCGTACGGCAGCCAGATCA 206
  |||||
85 AlaGluIlePheIleAspAlaAlaLysMetIleLeuHisProAspTyrAs 101
  |||||
207 CAGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 256
  |||||
101 PLYSTrPThrValAspAsnAspIleMetLeuIleLysLeuLysSerProA 118
  |||||
257 TGTCCGAGTCTGACACCATCCGAGCATCAGATTCCTGCTGCTGCTGCTGCT 306
  |||||
118 IatThrLeuAsnSerLysValSerThrIleProLeuProGlnTrpCysPro 134
  |||||
307 ACCGCGGGAACCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 356
  |||||
135 ThrIleGlyThrGlnCysLeuValSerGlyTrpGlyValLeuLysPheG1 151
  |||||
357 C...AGAATGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 403
  |||||
151 yPheGluSerProSerValLeuGlnCysLeuAspAlaProValLeuSerA 168
  |||||
404 AGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 453
  |||||
168 spSerValCysHisLysAlaIleTrpArgGlnIleThrAsnAsnMetPhe 184
  |||||
454 TGGCGCGCGGAGGCGAAGACAGAGACTCTGCAACGGTACTCTGG 503
  |||||
185 CysLeuGlnPheLeuGlnGlyLysAspSerCysGlnTrpAspSerG1 201
  |||||
504 GGGGCGGCTGATCTGCAACGGGCTACTGCAAGGCTTGTCTTTCGGA 553
  |||||
201 yGlyProValValCysAsnGlyLysValGlnGlyIleValSerTrpGlyA 218
  |||||
554 AAGCGCGGTGGCCAACTTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 603
  |||||
218 spGly...CysAlaLeuGlnGlyLysProGlyValIleThrLysValCys 233
  |||||
604 AATTCACGTAGTGATAGAGAAACCGTC 633
  |||||
234 AsnTrpLeuAsnTrpIleGlnIleThrVal 243
  |||||

```

```

seq_name: p1r2:S33772
seq_documentation_block:
  tissue kallikrein (EC 3.4.21.35) precursor, renal - crab-eating macaque
  C:Species: Macaca fascicularis (crab-eating macaque)
  C:Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 08-Sep-1997
  C:Accession: S33772
  R:Lin, F.K.; Lin, C.H.; Chou, C.C.; Chan, K.; Lu, H.S.; Bachelier, W.; Herrera, C.; J
  Biochim. Biophys. Acta 1173, 325-328, 1993
  A:Title: Molecular cloning and sequence analysis of the monkey and human tissue kall
  A:Reference number: S33772; MID:93305727
  A:Accession: S33772
  A:Status: preliminary
  A:Molecule type: mRNA
  A:Residues: 1-257 <LIN>
  A:Cross-references: GB:I10039; NID:9293140; PID:9293141
  A:Note: the authors translated the codon CGT for residue 17 as Ala, GCG for residue 1

```


OM of: US-09-030-606-175 to: SwissProt_37.* out_format: pfs
Date: Sep 25, 1999 4:46 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:

-MODEL=framer+np.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09030606/runat.24091999.171618.29883/app-query.fasta.1
-DB=SwissProt_37 -QFMT=fastan -SUFFIX=isp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -YGAPOP=10.000 -YGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsun62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR.SCORE=esscore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:

Query: US-09-030-606-175
Query length: 1167
Database: SwissProt_37.*
Database sequences: 77977
Database length: 28268293
Search time (sec): 155.430000

score_list:
Sequence Strd Orig ZScore EScore Len Documentation
SwissProt_37:SCCE_HUMAN + 530.50 551.14 1.5e-23 253 | P49862 homo sapiens (human).
SwissProt_37:KLK2_RAT + 478.00 497.87 1.5e-20 244 | P36375 rattus norvegicus (rat).
SwissProt_37:KLK3_MOUSE + 472.00 493.28 3.2e-20 261 | P00756 mus musculus (mouse).
SwissProt_37:TRYL_CHICK + 470.50 490.11 3.9e-20 248 | Q90627 gallus gallus (chicken).
SwissProt_37:RSS9_HUMAN + 470.50 490.23 3.9e-20 244 | Q92876 homo sapiens (human).
SwissProt_37:TRV2_RAT + 469.50 489.12 4.4e-20 247 | P08426 rattus norvegicus (rat).
SwissProt_37:TRV2_CHICK + 468.50 488.07 5.1e-20 248 | Q90628 gallus gallus (chicken).
SwissProt_37:TRV2_XENLA + 468.50 488.19 5.1e-20 244 | P70059 xenopus laevis (afrod).
SwissProt_37:TRV3_CHICK + 467.50 487.05 5.8e-20 248 | Q90629 gallus gallus (chicken).
SwissProt_37:KLK8_RAT + 463.00 482.10 1.0e-19 261 | P36374 rattus norvegicus (rat).
SwissProt_37:KLK8_RAT + 463.00 482.16 1.0e-19 259 | P36377 rattus norvegicus (rat).
SwissProt_37:TRVB_RAT + 462.00 481.50 1.2e-19 246 | P36382 rattus norvegicus (rat).
SwissProt_37:KLK1_MAFPA + 461.50 480.68 1.3e-19 257 | Q07276 macaca fascicularis (rat).
SwissProt_37:KLK7_RAT + 461.00 480.06 1.3e-19 261 | P36373 rattus norvegicus (rat).
SwissProt_37:KLK4_MOUSE + 461.00 480.06 1.3e-19 261 | P15946 mus musculus (mouse).
SwissProt_37:TRV1_CANFA + 459.50 478.95 1.6e-19 246 | P06871 canis familiaris (dog).
SwissProt_37:TRV1_RAT + 459.00 478.44 1.8e-19 246 | P36381 rattus norvegicus (rat).
SwissProt_37:KLK2_MOUSE + 458.00 477.00 2.0e-19 259 | P36368 mus musculus (mouse).
SwissProt_37:KLK2_MOUSE + 458.00 477.06 2.0e-19 259 | P15948 mus musculus (mouse).
SwissProt_37:TRV1_RAT + 457.00 476.38 2.3e-19 247 | P12788 rattus norvegicus (rat).
SwissProt_37:TRV1_BOVIN + 456.50 475.98 2.4e-19 243 | P00760 bos taurus (bovine).
SwissProt_37:KLK1_PAVHA + 456.00 475.05 2.6e-19 258 | Q28773 papio hamadryas (hama).
SwissProt_37:KLK9_MOUSE + 452.00 470.89 4.4e-19 261 | P15949 mus musculus (mouse).
SwissProt_37:KLK2_MOUSE + 450.00 468.85 5.7e-19 261 | P36369 mus musculus (mouse).
SwissProt_37:KLK2_MOUSE + 449.00 467.83 6.4e-19 261 | P20151 homo sapiens (human).
SwissProt_37:PROS_HUMAN + 449.00 467.83 6.4e-19 261 | P07288 homo sapiens (human).
SwissProt_37:TRV2_MOUSE + 448.50 467.74 6.9e-19 246 | P07286 mus musculus (mouse).
SwissProt_37:KLK2_RAT + 448.00 466.86 7.4e-19 259 | P00758 rattus norvegicus (rat).
SwissProt_37:KLK1_RAT + 447.00 465.79 8.4e-19 261 | P00759 rattus norvegicus (rat).
SwissProt_37:TRV1_XENLA + 446.00 464.77 9.5e-19 261 | P15947 mus musculus (mouse).
SwissProt_37:TRV1_XENLA + 445.00 464.26 1.1e-18 243 | P19799 xenopus laevis (afrod).
SwissProt_37:TRV1_RAT + 442.50 461.62 1.5e-18 246 | P00762 rattus norvegicus (rat).
SwissProt_37:TRV3_SALSA + 442.50 461.86 1.5e-18 238 | P35033 salmo salar (atlantic).
SwissProt_37:KLK3_MOUSE + 442.00 460.69 1.6e-18 261 | P15945 mus musculus (mouse).
SwissProt_37:TRV2_BOVIN + 441.50 460.57 1.7e-18 247 | Q29463 bos taurus (bovine).
SwissProt_37:TRV2_PIG + 440.50 460.03 2.0e-18 231 | P00761 sus scrofa (pig).
SwissProt_37:TRV2_CANFA + 438.50 457.52 2.6e-18 247 | P06872 canis familiaris (dog).
SwissProt_37:TRV2_SONAC + 438.00 457.54 2.7e-18 229 | P00764 squallus acanthias (sf).
SwissProt_37:KLK6_MOUSE + 437.00 455.60 3.1e-18 261 | P00755 mus musculus (mouse).
SwissProt_37:KLK1_HUMAN + 436.50 455.06 3.3e-18 262 | P06870 homo sapiens (human).
SwissProt_37:PROS_MACMU + 435.00 453.56 4.0e-18 261 | P36369 macaca mulatta (thesh).
SwissProt_37:TRV1_HUMAN + 434.50 453.44 4.3e-18 247 | P07477 homo sapiens (human).

seq_name: SwissProt_37:SCCE_HUMAN
seq_documentation_block:
ID SCCE_HUMAN STANDARD; PRT; 253 AA.
AC P49862;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE STRATUM CORNEUM CHYMOTRYPTIC ENZYME PRECURSOR (EC 3.4.21.-) (SCCE).
GN PRSS6 OR SCCE.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.
RC TISSUE=SKIN;
RX MEDLINE: 94308225.
RA HANSSON L., STROMQVIST M., BAECKMAN A., WALLBRANDT P., CARLSTEIN A.,
RA ESKELUND T.;
RT "Cloning, expression, and characterization of stratum corneum
RT chymotryptic enzyme. A skin-specific human serine proteinase.";
RL J. BIOL. CHEM. 269:19420-19426(1994).
RN [2]
RX CHARACTERIZATION.
RX MEDLINE: 95314630.
RA SKITT A., STROMQVIST M., ESKELUND T.;
RT "Primary substrate specificity of recombinant human stratum corneum
RT chymotryptic enzyme.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 211:586-589(1995).
CC -1- FUNCTION: MAY CATALYZE THE DEGRADATION OF INTERCELLULAR COHESIVE
CC STRUCTURES IN THE CORNIFIED LAYER OF THE SKIN IN THE CONTINUOUS
CC SHEDDING OF CELLS FROM THE SKIN SURFACE. SPECIFIC FOR AMINO ACID
CC RESIDUES WITH AROMATIC SIDE CHAINS IN THE P1 POSITION. SCCE
CC CLEAVES INSULIN A CHAIN AT 6-LEU-1-CYS-7, 16-TYR-1-LEU-17, 25-PHE-
CC 1-TYR-26, AND 26-TYR-1-THR-27. COULD PLAY A ROLE IN THE ACTIVATION
CC OF PRECURSORS TO INFLAMMATORY CYTOKINES.
CC -1- TISSUE SPECIFICITY: IT IS ABUNDANTLY EXPRESSED IN THE SKIN AND IS
CC EXPRESSED BY KERATINOCYTES IN THE EPIDERMIS. VERY LOW LEVELS ARE
CC ALSO SEEN IN THE BRAIN AND KIDNEY.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L33404; G532804; -;
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PIRAM: P00089; trypsin; 1.
DR HSSP: P00763; IDPO.
KW HYDROLASE; SERINE PROTEASE; ZMOGEN; GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 22
FT PROPEP 23 29
FT CHAIN 30 253
FT ACT_SITE 70 70
FT ACT_SITE 112 112
FT ACT_SITE 205 205
FT DISULFID 36 137
FT DISULFID 55 71
FT DISULFID 144 211
FT DISULFID 176 190
FT DISULFID 201 226
FT CARBOHYD 246 246
SQ SEQUENCE 253 AA; 27525 MW; 07FDB9F7 CRC32;

alignment_scores:

Quality: 530.50 Length: 214
 Ratio: 3.255 Gaps: 3
 Percent Similarity: 76.168 Percent Identity: 47.196

alignment_block:

US-09-030-606-175 x SCCE_HUMAN ..

Align seg 1/1 to: SCCE_HUMAN from: 1 to: 253

```

1 GCGAGCCCTGGAGCGGCGACATGTCATGAAAAAGATTGTCGTC 50
   ::::::::::::::::::::
40 SerHisProTprGlnValAlaLeuLeuSerGlySngInLeuHscYsgl 56
   ::::::::::::::::::::
56 YglYValLeuValAsnGlnArgTprValLeuThrAlaIahScYslsm 73
   ::::::::::::::::::::
101 AGACTCTACACCATGGGCTGGGCTGCACAGTCTGAGCGGACCA 150
   ::::::::::::::::::::
73 etAnGluTyrThrValHisLeuGlySerAspThrLeu..GlyAspArg 88
   ::::::::::::::::::::
151 GAGCCAGGAGCCAGATGTCGAGAGCCAGCTCCGTCGAGCAACCA 200
   ::::::::::::::::::::
89 Arg.....AlaGlnArgIleLysAlaSerLysSerPheArgHisProgl 103
   ::::::::::::::::::::
201 GTACACAGACTCTTCCTCGCTACAGACTCATGCTCATCAAGTTGACG 250
   ::::::::::::::::::::
103 YtyrSerThrInThrHisValAsnAspLeuMetLeuValLysLeuAsn 120
   ::::::::::::::::::::
251 AATCCGTTCGAGTCTGACACATCCGAGCAACAGATGTCGTCGAG 300
   ::::::::::::::::::::
120 erGlnAlaArgLeuSerSerMetValLysLysValArgLeuProSerArg 136
   ::::::::::::::::::::
301 TGCCCTACCGGGGAGACTCTGCTGCTGCTGCTGCTGCTGCTGCTG 350
   ::::::::::::::::::::
137 CysgluProProglYThrThrCysThrValSerGlyTprGlyThrThrTh 153
   ::::::::::::::::::::
351 GAAGGCG.....AGAATGCTACCGTCTGCTGCTGCTGCTGCTGCTG 394
   ::::::::::::::::::::
153 rSerProAspValThrPheProSerAspLeuMetCysValAspValLys 170
   ::::::::::::::::::::
395 TGCTGTCTGAGGAGTCTGACATGCTCTATGACCCGCTGACCAACC 444
   ::::::::::::::::::::
170 euIleSerProGlnAspCysThrLysValTyrLysAspLeuGlnAsn 186
   ::::::::::::::::::::
445 AGCATGTTCTGCGCGCGGCGGAGGCAAGACAGCAAGATCTCGCAACG 494
   ::::::::::::::::::::
187 serMetLeuCysAlaGlyIleProAspSerLysLysAsnAlaCysAsnGl 203
   ::::::::::::::::::::
495 TGACTCTGGGGGGGGCGCTGATCTGCAAGGCTACTGAGGGCGCTGTGT 544
   ::::::::::::::::::::
203 YAspSerIleGlyProLeuValCysArgGlyThrLeuGlnIleuValS 220
   ::::::::::::::::::::
545 CTTCGCAAAAGCCCGTGTGCGCAACTGTCGCGTCCAGGTGTACACC 594
   ::::::::::::::::::::
220 erTprGlyThrPheProCysGlyGlnProAsnAspProGlyValTyrThr 236
   ::::::::::::::::::::
595 AACCTGTGCAATTCATCTAGTGTGATAGAGAAACCGTCCAG 636
   ::::::::::::::::::::
237 GlnValCysLysPheThrLysThrIleAsnAspThrMetLys 250
   ::::::::::::::::::::

```

seq_name: SwissProt_37:KLK_RAT

seq_documentation_block:

ID KLK_RAT STANDARD; PRT; 244 AA.
 AC P36375;
 DT 01-JUN-1994 (REL. 29, CREATED)
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE GLANDULAR KALLIKREIN 10 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)

```

DE (T-KININOGENASE) (K10) (PROTEINASE B) (ENDOPEPTIDASE K) (FRAGMENT).
GN KLK10 OR KLK-10.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA.
OC RODENTIA; SCIROGNATHI; MURIDAE; MORINAE; RATTUS.
RN (1)
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-SUBMANDIBULAR GLAND, AND KIDNEY.
RX MEDLINE: 93041794.
RA MA J.-X., CHAO J., CHAO L.;
RT "Molecular cloning and characterization of rklk10, a cDNA encoding T-
RT kininogenase from rat submandibular gland and kidney.";
RL BIOCHEMISTRY 31:10922-10928(1992).
RN (2)
RP SEQUENCE OF 10-32; 95-124 AND 179-232.
RX MEDLINE: 91224135.
RA GUTMAN N., ELMOUHADH A., BRILLARD M., DE SORBIER B., GAUTHIER F.;
RT "Microheterogeneity of rat submaxillary gland kallikrein k10, a
RT member of the kallikrein family.";
RL EUR. J. BIOCHEM. 197:425-429(1991).
RN (3)
RP SEQUENCE OF 10-32 AND 97-133.
RC TISSUE-SUBMAXILLARY GLAND;
RX MEDLINE: 90153911.
RA XIONG W., CHEN L.-M., CHAO J.;
RT "Purification and characterization of a kallikrein-like
RT T-kininogenase.";
RL J. BIOL. CHEM. 265:2822-2827(1990).
RN (4)
RP SEQUENCE OF 10-32 AND 97-117.
RX MEDLINE: 88198057.
RA KATO H., NAKANISHI E., ENJOJOI K., HAYASHI I., OH-ISHI S., IWANAGA S.;
RT "Characterization of serine proteases isolated from rat
RT submaxillary gland: with special reference to the degradation of rat
RT kininogens by these enzymes.";
RL J. BIOCHEM. 102:1389-1404(1987).
RN (5)
RP FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
IN KININOGEN TO RELEASE LYS-BRADYKININ. THIS PROTEIN MAY BE
INVOLVED IN THE REGULATION OF RENAL FUNCTION.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-1-XAA BONDS IN
SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
KALLIDIN (LYSTL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
MET-1-XAA OR LEU-1-XAA.
CC -1- SUBUNIT: HETERODIMER OF A LIGHT CHAIN AND HEAVY CHAIN LINKED
BY A DISULFIDE BOND.
CC -1- TISSUE SPECIFICITY: KIDNEY AND SUBMANDIBULAR GLAND, WHERE IT IS
FOUND IN THE GRANULAR CONVOLUTED TUBULE AND STRIATED DUCT CELLS.
IT IS LIKELY THAT THE ENZYME IS MAINLY SYNTHESIZED IN THE GRANULAR
CONVOLUTED TUBULES AND THEN TRANSFERRED TO OTHER TISSUES BY
RELEASE INTO THE VASCULATURE OR INTERSTITIAL SPACE.
CC -1- PTM: PROBABLY N- AND O-GLYCOSYLATED. IT HAS CARBOXYMATE MOETIES
OF ALPHA-METHYL-D-MANNOSIDE AND N-ACETYL-D-GLUCOSAMINE GROUPS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S48142; G259431;
DR PIR: A35545; A35545;
DR PIR: B35545; B35545;
DR PIR: A44284; A44284;
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PRAM: P00089; TRYPSIN; 1.
DR HSSP: P00759; ITOM.
KW HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN; SIGNAL;
KW MULTIGENE FAMILY.

```


CC TRYPsin FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X01389; G53374; -

DR EMBL: X01788; G582429; -

DR EMBL: X01799; E73575; -

DR PIR: A00942; NMSG.

DR PDB: 1SGF; 27-MAY-98.

DR MGD: MGI:97322; NGFG.

DR PROSITE: PS00134; TRYPsin_HIS; 1.

DR PROSITE: PS00135; TRYPsin_SER; 1.

DR PFAM: PF00089; trypsin; 1.

KM HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN;

KM MULTIGENE FAMILY; ZMOGEN; SIGNAL; GROWTH FACTOR; 3D-STRUCTURE.

FT SIGNAL 1 18

FT PROPEP 19 24

FT CHAIN 25 261

FT CHAIN 25 261

FT CHAIN 112 261

FT ACT_SITE 65 65

FT ACT_SITE 120 120

FT ACT_SITE 213 213

FT DISULFID 31 173

FT DISULFID 50 66

FT DISULFID 152 219

FT DISULFID 184 198

FT DISULFID 209 234

FT CARBOHYD 102 102

FT DOMAIN 25 107

FT DOMAIN 112 261

FT DOMAIN 112 261

FT DOMAIN 165 261

FT CONFLICT 108 111

SEQ SEQUENCE 261 AA; 28998 MW; 7B9C5256 CRC32.

alignment_scores:

Quality: 472.00 Length: 228

Ratio: 2.969 Gaps: 4

Percent Similarity: 69.737 Percent Identity: 39.912

alignment_block:

US-09-030-606-175 x KLR3_MOUSE ..

Align seg 1/1 to: KLR3_MOUSE from: 1 to: 261

1 GCGGAGCCCTGCGAGCGGCGACGTGATGGAAGCAATGTTGCTC 50

35 serlnprtrprhlsvalalavalayrgrtyrtrhgrintyrlencysgl 51

51 GGGGCTGCTGTCGATCGGAGTGGGCTGTGACGCCGACAGTTCC 100

51 yglvalleuleusprroasnttrpvalletthralahlsctyrya 68

101 AGAAGCTCTACACCATCGGCTGGGCTGCACAGCTGTGAGGCCAGCA 150

68 spaprasntyrllyvaltrpleuglylysasnasnleuphelyasp... 83

151 GAGCCAGGAGCCAGATGTTGAGAGCCAGCTCTCCGTAAGGACCCAGA 200

84 Gluproserrallaglnhlsargrphvalserlysalatlelprohlsprogl 100

201 GTAAAGAGACTCTTGCTC 219

100 yphesnmsetserleuemetarglyshlsileargrphleuglutyraapt 117

220 ..GCTAAGACCTCATGCTCATCATGAGTGGACCAATCGTCCGAGTCT 267

117 yiserasnaspheulemetleuleuargleuserlysrprohlsaplethr 133

268 GACACCATCCGAGACATCATGCTGCTCGACAGTCCCTACCCGGGAA 317

134 Asphtrvallysrprolethrleuprothrglgluprolyslenglyse 150

318 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 367

150 rthncysleualaserlytrpqlserlietnr.....Prot 163

368 CCGTG.....CTGCATGCTGAGACGTGCTGCTGCTGCTGCTGCTGCT 399

163 hrlysrphneglnpethraspaspheuleuylcysvalasnleuylsleu 179

400 TCTGAGAGAGTCTGACATGCTGTATGACCCGCTGACACCCAGCAT 449

180 Proasncluaspcysalalysalahlslleglulysvalthrspalame 196

450 GTTCTGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 499

196 lencysalaglyglunetaspqlglylylsasprhncyslysglyasp 213

500 CTGGGGGGCCCTGATCTGCAAGGAGTCTGCAAGGAGGAGGAGGAGTCT 549

213 erglyglyproleuilecysaspqllylvalleuglnglylietnrsertrp 229

550 GGAAGACCCCGTGTGCGCACTTGGCGTCCAGAGTCTACACCAACT 599

230 Glylshtrpocysglygluproaspmetploglylvaltyrthrlyse 246

600 CTGCAATTCAGTGTGATGATGAGAAACCGTC 633

246 uasnlysrphetsertrpilleysasprhmet 257

seq_name: SwissProt_37:TRYL_CHICK

seq_documentation_block:

ID TRYL_CHICK STANDARD: PRT: 248 AA.

AC 090637;

DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

OS TRYPsinogen I-P1 PRECURSOR.

OS GALUS GALLUS (CHICKEN)

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;

OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.

OC [1]

RP SEQUENCE FROM N.A.

RC TISSUE-PANCREAS;

RX MEDLINE: 95251611.

RA WANG K., GAN L., LEE I., HOOD L.E.;

RT Isolation and characterization of the chicken trypsinogen gene

RL BIOCHEM. J. 307:471-479(1995).

CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-

CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.

CC -1- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE

CC LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPsin FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U15155; G603903; -

DR PROSITE: PS00134; TRYPsin_HIS; 1.

211 a l s e r t r p g l y a s n l l e p r o c g l y s e r l y s g l u l y s p r o g l y v a l t y r 227

7 CCCTGGCAGCGGCACCTGTCATGGAACGAATTGTTCC

7 CCCTGGCAGCGGCACCTGGTCATGGAAACGAATTGTTCTGCTCGGGCGT 56

```

37 ProTyrGlnValSerLeu...AsnAlaGlyTyrHisPheCysGlyGlySe 52
57 CQTGTGATCCGAGTGGTGTCTGACGCGCAGCTTTCGAGAACT 106
52 rLeuIleAsnSerGlnTyrValSerAlaIleHisCysTyrLysSer 69
107 CQTACACCATCGGCTGGCTGACAGTCTTGAGCGCCACCAAGCCA 156
69 rGleGlnValArgLeuGlyGlnHisAsnIleAspVal...ValGlnGly 84
157 GGGAGCCAGATGTGGAGCCAGCTCTCCGACGACCCAGAGTACAA 206
85 GlyGlnGlnPheIleAspAlaIleLysIleIleArgHisProSerTyrAs 101
207 CAGACTTGTGCTGCTAAGCAGCTCATGCTCATGAGTGGAGATCG 256
101 nAlaAsnThrPheAspAsnAspIleMetLeuIleLysLeuAsnSerPro 118
257 TGTCGAGCTGTGACACATCCGAGCATCAGCATGCTTCGAGTGCCT 306
118 lAlThrLeuAsnSerArgValSerThrValSerLeuProArgSerGly 134
307 ACCGGGGGAACTCTTGCCTGCTGTCGCTGGGCT...CTGCTGCG 350
135 SerSerGlyThrLysCysLeuValSerGlyTyrGlyAsnThrLeuSerSe 151
351 GAACGGCAGAAATGCTTACCGTCTGCTGCTGCTGCTGCTGCTGCT 400
151 rGlyThrAsnTyrProSerLeuLeuGlnCysLeuAspAlaProValLeu 168
401 CTGAGAGAGTGTGACATGCTATGACCCGCTGACACCCGAGCATG 450
168 eArgSerSerCysLysSerSerTyrProGlyLysIleThrSerAsnMet 184
451 TTTCGCGCGCGGAGGAGGAGCAGACAGACGCTCTGACAGCTGACT 500
185 PheCysLeuGlyPheLeuGlnGlyLysAspSerCysGlnGlyAspSe 201
501 TGGGGGGCGCGGATGTCAGAGGCTGACTGAGGCGCTGCTGCTGCTG 550
201 rGlyGlyProValValCysAsnGlyGlnLeuGlnIleValValSerTyr 218
551 GAAAGCCCGCTGCGCAACTGCGGCTGCGGCTGCTGCTGCTGCTGCT 600
218 lYTyGly...CysAlaGlnLysGlyLysProGlyValTyrThrLysVal 233
601 TGCAGATTCAGTGTGATGATGATGAGAAAACCGTC 633
234 CysAsnTyrValAsnTyrPheGlnGlnThrVal 244
seq_name: SwissProt_37:TRY2_CHICK
seq_documentation_block:
ID TRY2_CHICK STANDARD; PRT; 248 AA.
AC G90628;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE TRYPSINOGEN I-P38 PRECURSOR.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAFOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RX MEDLINE: 95251611.
RA WANG K.; GAN L.; LEE I.; HOOD L.E.;
RT Isolation and characterization of the chicken trypsinogen gene
family.
RL BIOCHEM. J. 307:471-479(1995).
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ANG-, LYS-.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- TISSUE SPECIFICITY: HIGH LEVELS ARE SEEN IN THE PANCREAS WHILE

```

```

CC LOWER LEVELS ARE FOUND IN THE LIVER, SPLEEN AND THYMUS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U15156; 6003905; -
CC PROSITE: PS00134; TRYPSIN_HIS. 1.
CC PROSITE: PS00135; TRYPSIN_SER. 1.
CC PRAM: PF00089; trypsin. 1.
CC DR HSP; P00763; IDPO.
CC KM HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
CC MULTIGENE FAMILY.
CC FT SIGNAL 1 15 BY SIMILARITY.
CC FT PROPEP 16 25 ACTIVATION PEPTIDE (BY SIMILARITY).
CC FT CHAIN 26 248 TRYPsin I-P38.
CC FT ACT_SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT DISULFID 32 162 BY SIMILARITY.
CC FT DISULFID 50 66 BY SIMILARITY.
CC FT DISULFID 134 235 BY SIMILARITY.
CC FT DISULFID 141 208 BY SIMILARITY.
CC FT DISULFID 173 187 BY SIMILARITY.
CC FT DISULFID 198 222 BY SIMILARITY.
CC FT SITE 196 196 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
CC SQ SEQUENCE 248 AA; 26087 MW; F64E0643 CNC32;

```

```

alignment_scores:
Quality: 468.50 Length: 216
Ratio: 2.910 Gaps: 5
Percent Similarity: 74.537 Percent Identity: 43.056

```

alignment_block:

US-09-030-606-175 x TRY2_CHICK

Align seg 1/1 to: TRY2_CHICK from: 1 to: 248

```

1 GCGCAGCCCTGGCAGCGGCGCAGTGTGATGGAAGAAAGATGCTGCTGCTC 50
36 AlaAlaProTyrGlnValSerLeu...AsnSerGlyTyrHisPheCysGly 51
51 GGGCTCTGTGTCATCGCAGTGGTGTCTGTCAGCGCCAGCATGTTCC 100
51 yGlySerLeuIleSerSerGlnTyrValLeuSerAlaIleHisCysTyrL 68
101 AGAATCTCTACACCATCGGCTGCGCTGACAGCTTGTGAGCGGACCA 150
68 ySerSerIleGlnValLysLeuGlyGlyTyrAsnLeuAlaIle...Gln 83
151 GAGCAGGAGGAGCAGATGATGAGGAGCGCCAGCTCTCCGACGACCA 200
84 AspGlySerGlnGlnThrIleSerSerSerLysValIleArgHisSerG 100
201 GTACAGAGACTCTTGTGCTGCTAAGCAGCTCATGCTCATGAGTTGACG 250
100 yTyrAsnAlaAsnThrLeuAsnAspIleMetLeuIleLysLeuSerL 117
251 AATCGGTGCGAGCTGACACATCCGAGCATCAGCATGAGTGTGCGAG 300
117 yAlaAlaThrLeuAsnSerTyrValAsnThrValProLeuProThrSer 133
301 TGCCCTACCGCGGGAACTCTGCTGCTGCTGCTGCTGCTGCTGCT 347
134 CysValThrAlaGlyThrThrCysLeuIleSerGlyTyrGlyAsnThrLe 150

```

```

348 GCGCAAGCGAGATG...CCTACCTGCTGCACTGCGTGAAGTGTGCG 394
||||| : : ||| ||||| : : : : : ||||| : : : : :
150 userSerGlySerLeuTyrProAspValLeuGlnCysLeuAsnAlaProV 167
395 TGTGTCTGTGAGANGTCTGCACTAAGCTCTATGACCCGCTGTACACCCC 444
||||| : : : : : ||||| : : : : : ||||| : : : : :
167 AlLeuSerSerSerGlnCysSerSerSerTyrProGlyArgIleThrSer 183
445 AGCATGTTCGCGCGGAGGAGGCAAGACAGAGACCTCCGCAAGCG 494
||||| : : : : : ||||| : : : : : ||||| : : : : :
184 AsnMetIleGlyIleGlyTyrLeuAsnGlyGlyLysAspSerCysGlnG 200
495 TGACTGTGGGGGGCCCTGATCTGCAACGGGCTACTTGCAGGCGCTGTGT 544
||||| : : : : : ||||| : : : : : ||||| : : : : :
200 yAspSerGlyGlyProValValCysAsnGlyGlnLeuGlnGlyPheValS 217
545 CTTTCGGAAGAACCCCGCTGTGGCAACTTGGCGTCCAGGTGTCTACACC 594
||||| : : : : : ||||| : : : : : ||||| : : : : :
217 eTrpGly...IleGlyCysAlaGlnGlySerGlyTyrProGlyValTyrThr 232
595 AACCTCTGCAATTCACCTAGTGATAGAGAAACGCTCCAGNCCAGT 642
||||| : : : : : ||||| : : : : : ||||| : : : : :
233 LysValCysAsnTyrValSerTyrIleTyrThrThrMetSerSerAsn 248

seq_name: SwissProt_37:TRY2_XENLA

seq_documentation_block:
ID TRY2_XENLA STANDARD: PRT: 244 AA.
AC P70059:
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE TRYPSINOGEN PRECURSOR.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPOIDEA; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RA WANG K., LITTLE L., GAN L., HOOD L.E.;
RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -1- CATALYTIC ACTIVITY: PREPERITIAL CLEAVAGE: ARG-, LYS-.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U72330; G1621633; -
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PFAM: PF00089; trypsin; 1.
DR HSSP: P00763; IDPO
KM HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZMOGEN; SIGNAL;
MULTIGENE FAMILY.
KW
FT SIGNAL 1 15
FT PROPEP 16 21 BY SIMILARITY.
FT CHAIN 22 244 ACTIVATION PEPTIDE (BY SIMILARITY).
FT ACT_SITE 61 61 TRYPSIN.
FT ACT_SITE 105 105 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 158 158 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 28 158 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 46 62 BY SIMILARITY.
FT DISULFID 130 231 BY SIMILARITY.
FT DISULFID 137 204 BY SIMILARITY.
FT DISULFID 169 183 BY SIMILARITY.
FT DISULFID 184 218 BY SIMILARITY.
FT SITE 192 192 REQUIRED FOR SPECIFICITY (BY SIMILARITY).

```

```

SQ SEQUENCE 244 AA; 26079 MW; 754B2E0E CRC32;
alignment_scores:
Quality: 468.50 Length: 216
Ratio: 2.947 Gaps: 4
Percent Similarity: 73.611 Percent Identity: 42.130
alignment_block:
US-09-030-606-175 x TRY2_XENLA ..
Align seg 1/1 to: TRY2_XENLA from: 1 to: 244
1 GCGCAGCCCTGGCGGCGGCACTGCTCATGAGAAACGAATGTGTGCTGTC 50
||| ||||| : : : : : ||||| : : : : :
32 AlaValProTyrGlnValSerLeu...AsnAlaGlyTyrHisPheCysG 47
47 yGlySerLeuIleAsnSerGlnTyrValValSerAlaIleHisCysTyrL 64
51 GGGCGTCTGTGTCATCCGCACTGGGTGCTGTACGCGCAGCATGTTTCC 100
||||| : : : : : ||||| : : : : : ||||| : : : : :
47 yGlySerLeuIleAsnSerGlnTyrValValSerAlaIleHisCysTyrL 64
101 AGACCTCTACACCATCGGCGTGGCGCTGCACAGTCTTGAGCCGACCA 150
||||| : : : : : ||||| : : : : : ||||| : : : : :
64 ySerArgIleGlnValArgLeuGlyGlnHisAsnIle...AlaLeuAsn 79
151 GAGCAGAGGAGCGACGATGGTGGAGCGGCGCTCTCGTACGCGACCCAGA 200
||| : : : : : ||||| : : : : : ||||| : : : : :
80 GlnGlyThrGlnGlnPheIleAspSerGlnLysValIleLysHisProAs 96
201 GTACACAGACATCTTCTGCTACAGACCTCATGCTCATCAAGTTGGACG 250
||||| : : : : : ||||| : : : : : ||||| : : : : :
96 nTyrAsnSerArgAsnLeuAspAsnAspIleMetLeuIleLysLeuSerT 113
251 AATCGCTGCGCGTGTGACACCATCGGACATCGACATCGCTTGTGGCAG 300
||||| : : : : : ||||| : : : : : ||||| : : : : :
113 hTrpIleArgLeuSerAlaAsnIleGlnSerValProLeuProSerAla 129
301 TGCCCTACCGCGGAGCACTTGGCTGCTGCTGCTGCTGCTGCTGCTGCT 344
||| : : : : : ||||| : : : : : ||||| : : : : :
130 CysAlaSerAlaGlyThrAsnLysLeuIleSerGlyTyrGlyAsnThrLe 146
345 GCTGGGGAAGCGGCAAGATCCCTACCGCTGCTGCTGCTGCTGCTGCTG 394
||||| : : : : : ||||| : : : : : ||||| : : : : :
146 userSerGlyThrAsnTyrProAspLeuLeuGlnCysLeuAsnAlaProI 163
395 TGTGTCTGTGAGANGTCTGCACTAAGCTCTATGACCCGCTGTACACCCC 444
||||| : : : : : ||||| : : : : : ||||| : : : : :
163 leuThrAspSerGlnCysSerAsnSerTyrProGlyIleThrLys 179
445 AGCATGTTCGCGCGGAGGAGGCAAGACAGAGACCTCCGCAAGCG 494
||||| : : : : : ||||| : : : : : ||||| : : : : :
180 AsnMetPheCysAlaGlyPheLeuAlaGlyGlyLysAspSerCysGlnG 196
495 TGACTGTGGGGGGCCCTGATCTGCAACGGGCTACTTGCAGGCGCTGTGT 544
||||| : : : : : ||||| : : : : : ||||| : : : : :
196 yAspSerGlyGlyProValValCysAsnGlyGlnLeuGlnGlyValValS 213
545 CTTTCGGAAGAACCCCGCTGTGGCAACTTGGCGTCCAGGTGTCTACACC 594
||||| : : : : : ||||| : : : : : ||||| : : : : :
213 eTrpGlyTyrGly...CysAlaGlnArgAsnTyrProGlyValTyrThr 228
595 AACCTCTGCAATTCACCTAGTGATAGAGAAACGCTCCAGNCCAGT 642
||||| : : : : : ||||| : : : : : ||||| : : : : :
229 LysValCysAsnPheValThrTyrIleGlnSerThrIleSerSerAsn 244

seq_name: SwissProt_37:TRY3_CHICK

seq_documentation_block:
ID TRY3_CHICK STANDARD: PRT: 248 AA.
AC Q90629:
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

```



```

207 CAGACTCTTGCTGCTAAAGCAGCCTCATGCTCATCAAGTTGGACATCG 256
      :  ::  |||  |||||:::|||||||:::|||||:
102 rserlIethrleuAnaasnaSPIlemtleuileuLysleuAlaSerAlav 119
      :  ::  |||  |||||:::|||||||:::|||||:
257 TGTCCGAGCTGCACACATCCGAGCATCAGCATGCTTCCGACATGCCCT 306
      :  ::  |||  |||||:::|||||||:::|||||:
119 aGIuLySerAlaasPIleGIInProIIeAlaLeuProSerSerCyAla 135
      :  ::  |||  |||||:::|||||||:::|||||:
307 ACCGGGGGAACTCTTGCTGTTCTGGCTGGGGGT...CTGCTGGCGAA 353
      |||||:::|||||:::|||||||:::|||||:
136 LysAlaGIyThrlucysleuileSerGIyTrpGIyAsnThrLeuSerAs 152
      :  ::  |||  |||||:::|||||||:::|||||:
354 CGGC...AGAATGCCCTACCGTGCCTGCTCAGTGAAGCTGTCGGTGT 400
      ||||  ::  |||  ::|||:::|||||:::|||||:
152 nGIyLyAsnTrpProGluLeuLeuGlnCysleuAsnAlaProIIeLeuS 169
      :  ::  |||  |||||:::|||||||:::|||||:
401 CTGAGAGANCTGTGCAGTAAGCTCTATGACCCGCTGTACCACCCAATG 450
      :  ::  |||  |||||:::|||||||:::|||||:
169 ezaSPglnglucysglnGlnAlaTrpProGlyAspIIeThrSerAsnMet 185
      :  ::  |||  |||||:::|||||||:::|||||:
451 TTCTGCGCGCGGGGAGGCGAAGACCAAGAGCTCTTCGACAGTGACTC 500
      :  ::  |||  |||||:::|||||||:::|||||:
186 lIeCyValGIyPheleuGlnIyGIyLysAspSerCySglnGIyAspSe 202
      :  ::  |||  |||||:::|||||||:::|||||:
501 TGGGGGGCCCGATCTGCACAGCGGTACTGTGAGGCGCTTGCTGTTTCG 550
      :  ::  |||  |||||:::|||||||:::|||||:
202 rGIyGIyProValIyAlCysAsnGIyGlnleuGlnIyLeuAlaSerTrpG 219
      :  ::  |||  |||||:::|||||||:::|||||:
551 GAAAGCCCCGTGTGGCCCACTTGGCGTGCAGGTGTCTACACCAATC 600
      :  ::  |||  |||||:::|||||||:::|||||:
219 lY...lIeGIyCyAlaLeuLysGIyTrpProGlyValTrpThrLysVal 234
      :  ::  |||  |||||:::|||||||:::|||||:
601 TGCCAATTCACAGTGGATAGAGAAAAACCGC 633
      :  ::  |||  |||||:::|||||||:::|||||:
235 CysAsnTrpValAspTrpIIeGlnGlnIuThrIle 245

seq_name: SwissProt_37:KlK8_RAT

seq_documentation_block:
ID KlK8_RAT STANDARD; PRT; 261 AA.
AC P36374;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE GLANDULAR KALLIKREIN 8, PROSTATIC PRECURSOR (EC 3.4.21.35) (TISSUE
DE KALLIKREIN) (PI KALLIKREIN) (RGR-8).
GN KLK8 OR KLK-8.
OS RATTUS NORVEGICUS (RAT).
OC EUMAROTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCURIONATHT; MORIDAE; MORINAE; RATTUS.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE; 89352606.
RA BRADY J.M., WINES D.R., MACDONALD R.J.;
RT "Expression of two kallikrein gene family members in the rat
RT prostate."
RL BIOCHEMISTRY 28:5203-5210(1989).
CC -1- FUNCTION: GLANDULAR KALLIKREIN CLEAVE MET-LYS AND ARG-SER BONDS
CC -1- KININOMEN TO RELEASE LYS-BRADYKININ.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-|-XAA BONDS IN
CC SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
CC KALLIDIN (LYSYL-BRADYKININ) FROM KININOMEN INVOLVES HYDROLYSIS OF
CC MET-|-XAA OR LEU-|-XAA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```



```

7  CCCTGGAGGGGGGCTGCTCATGGAAGAAACGAATGTTCTGCGGGCT 56
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
37  Protyrinalserleuasnalaiglyserhisile...Cysglylyse 52
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
57  CCTGGTCATCCGAGGAGGCTGCTGTCAGCCGACACGTTTCCAGAACT 106
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
52  rleuethraspargintrpvalleuseralaalahlscystyrhispro 69
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
107  CCTACACCATGGGCTGGCGCTGACAGTCTTGGAGCGCCAGCAAGAGCA 156
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
69  lndeuainvalargleuglylunhisasnte...Tyrghilileugly 84
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
157  GGGAGCAGATGTGTGAGGCCAGGCTCTCCGTAGGACCCAGAGTACAA 206
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85  Aladuglnphelileaspalaalalysmetileuhsiproasptyras 101
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207  CAGACTCTGTGCTGCTAACGACCTCATGCTATCATAGTTGACAAATCCG 256
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101  pLysTrpThrValaspasnaspillemleuileuylsleuylsSerProA 118
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
257  TGTCGAGTGTGACACCATCCGAGCATCAGCATGTTGTCGAGTGCCT 306
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
118  latrpheuasnerlyvalserthrileproleproglintyrCyspro 134
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
307  ACCGCGGGAACTTGTGCTGTTGCTGCTGGGCTGTGCTGGGAAACG 356
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
135  ThrIaaglythrIuGlylsleuValserglyTyrpIyValleuylsPhegl 151
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
357  C...AGATGCTTACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 403
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151  yPhegluSerProserValserThrileproleproglintyrCyspro 168
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
404  AGANGTCTGAGTAAGCTGTATACCCGCTGTACACCCAGCATGTC 453
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
168  spserValCysHisIysIasIatyrProargIlnIleThrAsnMethe 184
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
454  TGGCCGCGGGAGGAGGACAGACAGAGACTCTGCAACGGTACTGG 503
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185  CysleuglyPheleuglyIyIyIysaspserCysglntyrAspsergl 201
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
504  GGGGCGCCCTGATCTGCAACGGTACTGCAAGGCTTGTGCTTTCGAA 553
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201  yglIprovalValCysasnIyIyIvalIghIyIleValsertrpIy 218
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
554  AAGCGCGGTGGCCAACTTGGCGTGCAGAGTGTATACCAACCTGTC 603
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218  spgly...CysalaleuGllylserProgllyValtyrThrlyValcys 233
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
604  AATTCAGTGAAGTGAAGAGAGAAACCGTC 633
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234  AsntyrlleuasntyrpIleGlInIghIlnIthryal 243
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: Swisprot_37:KRL1_MACFA

seq_documentation_block:
ID KRL1_MACFA STANDARD; PRT; 257 AA.

```

AC 007276;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE GLANDULAR KALLIKREIN 1 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)
DE (KIDNEY/PANCREAS/SALIVARY GLAND KALLIKREIN).
GN KRL1.
OS MACACA FASCICULARIS (CRAB EATING MACAQUE) (CYNOMOLGUS MONKEY).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; MACACA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93305727.
RA LIN F.K., LIN C.H., CHOU C., CHEN K., LU H.S., BACHFELLER B.,
RA HERRERA C., JONES T., CHAO J., CHAO L.
RT "Molecular cloning and sequence analysis of the monkey and human
RT tissue kallikrein genes."

```

```

RL BIOCHIM. BIOPHYS. ACTA 1173:325-328(1993).
CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -1- CATALYTIC ACTIVITY: PREPERENTIAL CLEAVAGE OF ARG-1-XAA BONDS IN
CC SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
CC KALLIDIN (LYSL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC MET-1-XAA OR LEU-1-XAA.
CC -1- SIMILARITY: BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC TRYPsin FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
CC DR EMBL; L10039; G293141; -.
CC DR PIR; S33772; S33772.
CC DR PROSITE; PS00134; TRYPsin_HIS; 1.
CC DR PROSITE; PS00135; TRYPsin_SER; 1.
CC DR PFAM; PF00089; trypsin; 1.
CC DR HSP; P00752; IHA.
CC KM HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN; PANCREAS;
CC ZMOGEN; SIGNAL.
CC FT SIGNAL 1 18 BY SIMILARITY.
CC FT PROPEP 19 24 ACTIVATION PEPTIDE (PROBABLE).
CC FT CHAIN 25 257 GLANDULAR KALLIKREIN 1.
CC FT ACT_SITE 62 62 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 116 116 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 209 209 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT CARBOHYD 90 90 POTENTIAL.
CC FT CARBOHYD 99 99 POTENTIAL.
CC FT CARBOHYD 101 101 POTENTIAL.
CC FT CARBOHYD 105 105 POTENTIAL.
CC FT CARBOHYD 160 160 POTENTIAL.
CC FT CARBOHYD 162 162 POTENTIAL.
CC FT DISULFID 31 169 BY SIMILARITY.
CC FT DISULFID 47 63 BY SIMILARITY.
CC FT DISULFID 148 215 BY SIMILARITY.
CC FT DISULFID 180 194 BY SIMILARITY.
CC FT DISULFID 205 230 BY SIMILARITY.
CC SQ SEQUENCE 257 AA; 28237 MW; 82A709DC CRC32;

```

alignment_scores:

Quality: 461.50 Length: 223
Ratio: 2.977 Gaps: 3
Percent Similarity: 69.507 Percent Identity: 40.359

alignment_block:

us-09-030-606-175 x KRL1_MACFA ..

Align seg 1/1 to: KRL1_MACFA from: 1 to: 257

```

1  GCGGAGCGCTGGGAGGCGGCTGTCATGGAAGAAACGAATGTTCTGCTC 50
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
32  SerIlnProtrpIghIlnAlaIaleuTyrlHisPheSerThrPheGlnCysgl 48
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51  GGGGCTGCTGTCATCCGAGTGGTGTGCTGTCAGCCGACACGTTTCC 100
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48  yglIyIleuValHisPheGlnItrpValleuThrAlaAlaHisCysIles 65
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101  AGAACTCTTACACATGGGCTGGGCTGTCAGACGCTTGGAGCCGAGCAA 150
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
65  eraSpasnTyrglnleuTrpIleuIyIargHisasnleu...Pheasp 80
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151  GAGCCAGGAGCAGATGTGTGAGGCCAGGCTCTCCGTAGGACCCAGAGA 200
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81  GluaspThrIaaglnPheValHisValSerGluSerPheProHisProgl 97
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201  GTACACAGACTCTTGTGCTGCTAAC..... 225

```

```

97 yphenasmetSerLeuLeuLysAsnHisThrArgGlnAlaAspArgLys 114
   ::||| ||||| |||
226 ....GACCTCATGCTCATGAGTGGAGCAATCCGTGCCAGTCTGC 270
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
114 erHisAspMetLeuLeuArgLeuThrGlnProAlaGluLeuThrAsp 130
271 ACCATCCGGAGCATGAGCATGCTTCGACAGTCCCTACCGCGGGAATC 320
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
131 AlaValGlnValAlaGluLeuProThrGlnGluProGluValGlySerTh 147
321 TTGCTCTGTTCTGCTGGCTGCTCTG.....CGAAGCGCAAAATC 364
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
147 rCysLeuAlaSerGlyTrpGlySerIleGluProGluAsnPheserPhep 164
365 CTACCGTCTGCTGCTGAGCAAGTGGTGGTGTCTGAGAGANTCTGC 414
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
164 rAspAspLeuGlnCysValAspLeuGluLeuLeuProAsnAspLeuLys 180
415 AGTAAGCTATGACCCGCTGACACCCACAGCATTTCTCGCCGCGG 464
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 AlaLysAlaHisThrGlnLysValThrGluPheMetLeuCysAlaGlyH 197
465 AGGCAAGACCAAGAGACTCTGACAGGTCATCTGGGGGGCCCTGA 514
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
197 sLeuGluLeuGlyLysAspThrCysValGlyAspSerGlyGlyProLeu 214
515 TCTGCAAGGGTACTTGCAGGCGCTGTGTCTTGGAAACCCCGGT 564
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
214 hrCysAspGlyValLeuGlnGlyValThrSerTrpGlyTrpIleProCys 230
565 GGCACACTGGCGTCGACGTCCTACACCAACTCTGCAATTCATCA 614
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
231 GlySerProAsnLysProAlaValPheValLysValLeuSerTyVally 247
615 GTGATAGAGAAACCGTC 633
   ::||| ||||| |||||
247 StrpIleGluAspThrIle 253

```

seq_name: SwissProt_37:KLK7_RAT

seq_documentation_block:

```

ID KLK7_RAT STANDARD; PRT; 261 AA.
AC P36373;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE GLANDULAR KALLIKREIN 7, SUBMANDIBULAR/RENAL PRECURSOR (EC 3.4.21.35)
DE (TISSUE KALLIKREIN) (RGK-7) (RSKG-7) (ESTERASE B) (PROTEINASE A).
GN KLK7 OR KLK-7.
OS RATIUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCOTROGNATHI; MORIDAE; MORINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89088074.
RA CHEN Y.-P., CHAO J., CHAO L.;
RT "Molecular cloning and characterization of two rat renal kallikrein
RT genes."
RL BIOCHEMISTRY 27:7189-7196(1988).
RN [2]
RP SEQUENCE OF 25-75.
RX TISSUE-SUBMAXILLARY GLAND;
RX MEDLINE: 88198057.
RA KATO H., NAKANISHI E., ENJOYI K., HAYASHI I., OH-ISHI S., IWANAGA S.;
RT "Characterization of serine proteinases isolated from rat
RT submaxillary gland: with special reference to the degradation of rat
RT kininogens by these enzymes."
RL J. BIOCHEM. 102:1389-1404(1987).
CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-1-XAA BONDS IN
CC SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE

```

```

CC KALLIDIN (LYSL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC MET-1-XAA OR LEU-1-XAA.
CC -1- TISSUE SPECIFICITY: KIDNEY AND SUBMANDIBULAR GLAND.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19647; G205000; -
DR PIR: A31136; A31136.
DR PIR: B41429; B41429.
DR PROSITE: PS00134; TRYPsin_HIS; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
DR PIR: P00089; trypsin; 1.
DR HSSP: P00759; 1TON.
KW HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN;
KW MULTIGENE FAMILY; ZMOGEN; SIGNAL.
FT SIGNAL 1 18
FT PROPEP 19 24
FT CHAIN 25 261
FT ACT_SITE 65 65
FT ACT_SITE 120 120
FT ACT_SITE 213 213
FT DISULFID 31 173
FT DISULFID 50 66
FT DISULFID 152 219
FT DISULFID 184 198
FT DISULFID 209 234
FT CARBOHYD 108 108
FT CONFLICT 35 35
FT CONFLICT 46 46
FT SEQUENCE 261 AA; 28972 MW; 27865842 CRC32;

```

alignment_scores: Quality: 461.00 Length: 225
Ratio: 2.918 Gaps: 3
Percent Similarity: 70.222 Percent Identity: 39.556

alignment_block:

US-09-030-606-175 x KLK7_RAT ..

Align seg 1/1 to: KLK7_RAT from: 1 to: 261

```

1 GCGAGACCCCTGGAGCGGCGCATGATGAAACCAATGTTCTGCTC 50
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
35 SerGlnProTrpGlnValAlaLeuLysSerPheThrLysTyLeuGlySgl 51
51 GGGGCTCTGGTGCATCCGACAGTGGTGTCTGACCGCACATGTTCC 100
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 yGlyValLeuLeuAspProSerTrpValIleThrAlaHisLysSers 68
101 AGAATCTCTACACCATCGGCTGGCTGACAGCTTGTAGCGGACCAA 150
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68 erAsnAsnTyGlnValTrpLeuGlyArgAsnAsnLeuGluAsp... 83
151 GAGCAGAGGAGCCAGATGCTGAGCGCACGCTCCGACGACCCAGA 200
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 GluProPheAlaGlnHisArgLeuValSerGlnSerPheProHisProAs 100
201 GTACAACAGACTCTTGCTC..... 219
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
100 pTyLysProPheLeuMetArgAsnHisThrArgLysProGlyLysAspSph 117
220 ..GCTAAGACCTCATGCTCATGAGTGGACGAAATCCGTGCTCGAGCT 267
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 IsSerAsnAspLeuMetLeuLeuHisLeuSerGlnProAlaAspIleThr 133

```


This Page Blank (uspto)


```

175  erg1UG1nHrthCysarGdeuLeuTYAsPrOValTYhns1euseSerMet 191
451  TTCTGCGCGCGGCGGAGGCGACAGACAGACTCTGCAACGGTACTC 500
|||||
192  Phcysa1agi1yGLyGLnAsp1nLyAspSerCysasnG1yAspse 208
501  TGGGGGGGCGCCGAGCTGCTCAACGGGTACTTGCAGGGCGCTTGCTTTCG 550
|||||
208  rclYslYPrO1eVal1CysasnAlGSerLeuGlnG1yValSerMetG 225
551  GAAAGCCCGCGTGCGCAACTGGCGGTGCGAGGTCTACACCAACTC 600
|||||
225  lyGlnG1yLyscysG1yInPrOg1yIleProSerValTYThAsn1eu 241
601  TGCAAATTCAGTGAAGTGATAGAGAAACCGTCCAGNCCACT 642
|||||
242  CysLysPheThrAsnTTrp1IleGlnThr1IleGlnThrAsn 255
seq_name: sp_rodent.063274

seq_documentation_block:
ID 063274 PREIMINARY; PRT; 235 AA.
AC 063274;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, last annotation update)
DE KALIKREIN (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia
NC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CC [1]
PP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=KIDNEY;
RA Z1WT2 C.B., MA J.X., CHAO J., CHAO L.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: I33839; AAA58781.1; -.
DR PFM: PF00089; trypsin. 1.
FT NON_TER 1
SQ SEQUENCE 235 AA; 26226 MW; E6DE5AD8 CRC32;

```

```
alignment_scores:
  Quality: 486.00      Length: 2277
  Ratio: 3.019         Gaps: 5
Percent Similarity: 70.925 Percent Identity: 42.291
```

alignment_block:

US-09-030-606-175 x Q63274

Align seg 1/1 to: Q63274 from: 1 to: 235

```

1 GCCACCCTGGCAGGCGGCATGTGTCATGAGAAACGAATTGTCCTC 50
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
11 SerGlnProTrrpLlnValAlaValIle.....AsnArgTyrLeuGlySgl 25

51 GGGCGTCTGTGCATCCGCAGATGGGTGCTGTCAACCCGACACATGTTTCC 100
  |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
25 yglValLeuIleAspProSerTrrpValIleIleAlaIleAsnIleGlySyrS 42

101 AGACATCTTACACATCGGGGCTGGGCTCCGACATCTTGGAGCGGACCA 150
  ::::: ||| :::: ||||| |||||::: |||
42 ehtIstYrYrIlnValLeuIleuGlyArgHisAsnLeuPheGluAsp... 57

151 GACCCAGGAGGACAGATGTGTGAGGACGACCTCTCCGACGACCCAGA 200
  |||||:::|||||:::|||||:::||||| ||| |||||:::
58 GluProPheAlaGlnIlyrArgPheValSerGlnSerPheProHisProAs 74

201 GTACACACAGACTTGTCTC..... 219
  |||||:::|||||:::|||||:::|||||:::|||||:::
74 pTyArgProPheLeuMetArgAsnHisIsthrArgIleIlnHrgIlyrAspT 91

220 ..GCTAACGACCTTCATGCTCATCAAGTGTGAGCAATCCGTGTCGAGTCT 267
  :::::|||||:::|||||:::|||||:::|||||:::|||||:::

```

```

91  yrSerAsnAspLeuMetLeuIleuHisIleuSerGluProIleAspIleThr 107
268 GACACCAATCCGGAGCATATAGCATTTGCTTCGACAGTCCCTACCGGGGAA 317
111 ..... ||| ..... ||| ..... ||| ..... |||
108 AspGlyValValIleAspLeuProThrGluIleuProGlyValGlyIse 124
318 SCTCTGCTCGCTGCTGCTGGCTGGGT.....CTGCTGGCGAAGC 355
124 rThrcysLeuValSerGlyTrpGlySerThrLysProIleuLeuSer... 139
356 GCAGATACCTTACCGTGTCTGCATCTGCGTGAACGTGGTGGTGTCTAG 405
140 .....||| .....|||.....|||.....|||.....|||
140 ..GluLeuProAspAspLeuGlnCysValAsnIleAspLeuLeuSerAsn 155
406 GANGTCTCAGTAAGCTCTATGACCCGGTACACCCGACAGATTTCTG 455
156 GlnLysCysIleGluAlaTyrAlaGlyTrpLysValThrAspLeuMetLeu 172
456 CGCCGGCGAGAGGCGAAGACAGAAAGCATCTCCGACAGGTACTGTGGG 505
172 salIleLysIleuGlnGluGlyLysAspAlaCysAsnGluYaspreGlyG 189
506 GCGCCCTATCTGCAACGGGTACTTGCAGGGCCTTGCTGTTCGGAAA 555
189 IyrProLeuIleLysAspLysValIleuGlnGlyLeuThrSerTrpGlySer 205
556 GCGCCGCTTGCGCAATTGGCGGTGCGAGTGTACAGTCTGCAACATCTGCA 605
1206 ValProGlyGluGlyProHisAsnProGlyLeuThrLysIleIleLys 222
606 ATTCACTAGTGGATAGAGAAACCGTCCAG 636
222 spherhSerTrpIleLysGluValMetLys 232

```

seq_name: sp_human:075837

seq_documentation_block:

ID	075837	PRELIMINARY;	PRT;	282	AA
----	--------	--------------	------	-----	----

DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE SERINE PROTEASE (Tlsp).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
[1]
RN
RP
RX MEDLINE; 98438738.
RA YOSHIDA S., TANGUCHI M., SUMOMO T., OKA T., HE X.P., SHIOSKA S.,
RT "cDNA cloning and expression of a novel serine protease, Tlsp.".
RL Biochem. Biophys. Acta 1399:225-228(1998).
DR EMBL; AB012917; BA833604.1; --
PR PPM; PF00089; trypsin.1.
FM Protease; Serine protease.
KM CHAIN 54 282 Tlsp.
SQ SEQUENCE 282 AA; 31059 MW; E9YE294C CRC32;

alignment_scores:

Quality:	485.50	Length:	218
----------	--------	---------	-----

Percent Similarity: 72.936 Percent Identity: 41.284

alignment_block:

US-09-030-606-175 x 075837 .

Align seg 1/1 to: 075837 from: 1 to: 282

1 GCGCAGCCCTGGCAGCGCGCACTGGTCATGAGMAAACGAATTGCTGCTC 500
 :::::::::::::::::::::::::::: :::::::::::::::::::::
 64 SerGlnProTyrGlnAlaAlaLeuPheGlnLysThrArgLeuLeuCysGI 800

```

51 GGGCGCTGCTGTCATCCGACGAGTGGTCTGTACGCCGACACTGTTC 100
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
80 yAlaThrLeuIleAlaProArgTrpLeuThrAlaAlaHisCysLeu 97
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
101 AGAACCTCTACACATCGGGCTGGGCTGCACAGTGTAGAGCCGACAA 150
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
97 ySPArGTrYrIleValHisLeuGlyGlnHisAsnLeuGln...LysGln 112
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
151 GAGCCAGGAGCCAGATGCTGGAGGCGACCTCCGTCAGGCGACCCACA 200
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
113 GluGlyCysGluGlnThrArgThrAlaThrGluSerPheProHisPro 129
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
201 GTACACAGACTCTTG.....CTCCGTACAGACTCATGCTCA 238
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
129 yPheAsnAsnSerLeuProAsnLysAspHisArgAsnAspIleMetLeu 146
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
239 TCAGATTGGACGAATCCGTGCTCCGAGCTGACACCATCCGAGACATCAG 288
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
146 aLlyMetAlaSerProValSerIleThrPrAlaValArgProLeuThr 162
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
289 ATTGCTTGCACTGGCCCTACCGGGGAGACTCTTGCTCTGTTCGCTG 338
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
163 LeuSerSerArgCysValThrAlaGlyThrSerCysLeuIleSerCylTr 179
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
339 GGGTCTGTGCGCAACGCG.....AGAACTGCTACCGCTGTCACCTGCG 382
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
179 polySerThrSerSerProGlnLeuArgLeuProHisThrLeuAlaGlySA 196
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
383 TGACAGTGTGCTGTGTGTGAGANGTCTGCATAGCTCTATGACCCG 432
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
196 laAsnIleThrIleGlnHisGlnLysCysGlnAsnAlaTrpProGly 212
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
433 CTGTACACACCCAGCATGTCTTCGCGCGCGGAGGCGACAGACAGAA 482
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
213 AsnIleThrAspThrMetValCysAlaSerValGlnGlnGlyCylLysAs 229
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
483 CTCTGCAACGCTGACTGTGGGGGCGCCCTGATGTGCACAGGGTACTTC 532
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
229 pSerYsGlnGlyAspSerGlyCylProLeuValCysAsnGlnSerLeuG 246
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
533 AGGCGCTGTGTCTTTGGAAAGCCCGCTGTGCCCACTTGCGCTGCCA 582
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
246 lngIylIleSerTrpGlyGlnAspProCysAlaIleThrArgLysPro 262
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
583 GGTGCTTACACCACTCTGCAATTCACATGAGTGAAGAAAACCT 632
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
263 GlyValIlyThrLysValCysLysTyValAspTrpIleGlnGluThrMe 279
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
633 CCAG 636
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
279 tlys 280
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
seq_name: sp.fodent:063275
seq_documentation_block:
ID 063275 PRELIMINARY; PRT: 239 AA.
AC 063275:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE KALLIKREIN (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-SUBMANDIBULAR GLAND;
RA ZINTZ C.B., MA J.X., CHAO J., CHAO L.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: L33840; AAA58782.1;
DR PFM: PF00089; trypsin.1.
FT NON_TER 1
SEQUENCE 239 AA; 26382 MW; 0609E3E6 CRC32;

```

```

alignment_scores:
  Quality: 477.00      Length: 224
  Ratio: 2.944
  Percent Similarity: 72.321      Percent Identity: 41.518
alignment_block:
US-09-030-606-175 x 063275 ..
Align seg 1/1 to: 063275 from: 1 to: 239
1 GCGAGCCCTGCGACAGCGCGACCTGTCATGGAAGAAATGTTCTGCTC 50
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
19 SerGlnProTrpGlnValAlaValIle.....AsnGluAspLeuCysG 33
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
51 GGGCGCTGCTGTCATCCGACAGTGGGTGCTGTACGCCGACACTGTTC 100
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
33 yGlyValLeuIleAspProSerTrpValIleThrAlaAlaHisCysTr 50
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
101 AGAACTCTACACCATCGGGCTGGGCTGCACAGTGTGAGCCGACCAA 150
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
50 eRAspAsnTrpHisValLeuLeuGlyGlnAsnAsnLeuSerGluAspVal 66
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
151 GAGCCAGGAGCAGATGCTGAGGCGCAGCCTCCGTCAGCGACCCAGCA 200
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
67 Gln.....HisArgLeuVal.....SerGlnSerPheArgHisProAs 79
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
201 GTACACAGACTCTTGCTC.....G 220
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
79 pTyIysProPheLeuMetArgAsnHisThrArgLysProLysAspTr 96
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
221 CTACAGACTCATGCTCATCAAGTTGACAGAAATCCGTGCTCCGACTGAC 270
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
96 eRAspAspLeuMetLeuLeuHisLeuSerGluProAlaAspIleThrAsp 112
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
271 AACATCCGCGAGCATGATGCTTCGAGTGCAGTCCGCGCGGGAATC 320
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
113 GlyValLysValIleAspLeuProThrLysGluProLysValGlySerTh 129
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
321 TTGCTCTGTTCTGCTGGGTGCTGCTG.....GCGAACGCGCAATGC 364
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
129 rCysLeuValSerGlyTrpGlySerThrAsnProSerGluTrpGluPheP 146
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
365 CTACCGTCTGCTGACTGCTGTAACGTGTGCTGTGTGAGGANGTCTGC 414
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
146 roAspAspLeuGlnCysValAsnIleHisLeuLeuSerAsnGluLysCys 162
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
415 AGTAAGCTCTATGACCCGCTGTACACCCCGCAATGTTCTGCGCGCGG 464
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
163 IleLysAlaTrpTrpLysGluLysValThrAspLeuMetLeuCysAlaGly 179
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
465 AGGCAAGACCCAGAGACTCTGCAACGCTGACTGTGGGGGCGCCCTGA 514
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
179 uLeuGlnGlyCylLysAspThrCysArgGlyAspSerGlyCylProLeu 196
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
515 TCTGCAACGGGTACTGCGAGGCGCTGTGTCTTTGGAAAGCCCGCTG 564
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
196 leCysAspGlyValLeuGlnGlyIleThrSerTrpGlySerValProCys 212
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
565 GGGCAACTGGCGGCGCGCTGCTACACCACTGCAATTCACATGCA 614
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
213 GlyLupProAsnLysProGlyIleTrpThrLysLeuIleLysPheThrSe 229
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
615 GTGATAGAGAAACCGTCAG 636
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
229 rTrpLeuLysGluValMetLys 236
   :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::
seq_name: sp.mammal:Q29474
seq_documentation_block:
ID Q29474 PRELIMINARY; PRT: 261 AA.
AC Q29474;

```


RP SEQUENCE FROM N.A.
 RC TISSUE-PANCREAS;
 RA DOUGLAS S.E., GALLANT J.W.,
 RT "Isolation of cDNAs for trypsinogen from the winter flounder,
 RT pleuronectes americanus."
 RL J. Mar. Biotechnol. 0:0-0(1998).
 DR EMBL: AF012462; AAC32751.1;
 DR PFAM: PF00089; trypsin; 1.
 KW Signal; Hydrolase.
 FT SIGNAL 1 15 POTENTIAL.
 FT CHAIN 23 250 TRYPSINOGEN 1.
 SQ SEQUENCE 250 AA; 27466 MW; D387394D CRC32;

alignment_scores:
 Quality: 448.50 Length: 217
 Ratio: 2.875 Gaps: 5
 Percent Similarity: 71.889 Percent Identity: 42.396

alignment_block:
 US-09-030-606-175 x 093265 ..

Align seg 1/1 to: 093265 from: 1 to: 250

```

1 GCGCAGCCCTGGCAGCGGCGACATGTCATGGAACGAAATGTTCTGCTC 50
   ::::::::::::::::::::
33 SerArgProTyrMetAlaSerLeuAsnTyrGlyTrpHis...PheCysG1 48
51 GGGCGCTCTGTGATCCGACAGTGGGTCTCTACGCCGACATGTTTCC 100
   ::::::::::::::::::::
48 YglValLeuIleAsnGlnTyrValLeuSerValAlaHisCysTrpT 65
101 AGACTCTCTACACC....ATCGGGCTGGGCGCTGCACAGTCTGAGGCC 144
   ||| ::::::::::::::::::::
65 YrAsnProTyrAlaMetGlnValMetLeuGlyGlnHisAsnLeuArgVal 81
145 GACCAAGAGCCGAGGAGCAGATGGTGGAGCGCCACTCTCCGTACGGCA 194
   ||| ::::::::::::::::::::
82 ...PheGlnGlyThrGlnLeuMetLysThrAspThrIleIleTrpH1 97
195 CCCAGATACACAGACTCTTGTCTGCTACGACCTCATGCTATCAAGT 244
   ||| ::::::::::::::::::::
97 sProAsnTyrAspTyrGlnThrLeuAspPheAsnIleMetLeuIleYsL 114
245 TGAGAGATCGTGGTGGAGTCTGACACATCCGAGACATGATTCCT 294
   ||| ::::::::::::::::::::
114 eutYrHisProValGlnValThrGlnAlaValAlaProIleSerLeuPro 130
295 TCGCAGTCCCTACCGCGGGGAGACTCTTCTCTGCTGCTGGGCTCT 344
   ||| ::::::::::::::::::::
131 SerSerCysProValGlyGlyThrProCysSerValSerGlyTrpGlyAs 147
345 GCTGGCGCAAGCGC.....AGATGCTTACCGTGTGCACTGCG 382
   ||| ::::::::::::::::::::
147 nThrAlaArgAspGlyAspAspValTyrMetProThrLeuLeuGlnCysM 164
383 TGAACGTGTGGTGTGTCAGAGANGTCTGACAGTAAGCTATGACCCG 432
   ::::::::::::::::::::
164 eCAspValProIleLeuAsnGlnGlnCysMetLysSerTyrProGly 180
433 CTGTACACCCAGACATGTTCTGCGCCGCGAGGCGAGGACAGCAAGAA 482
   ::::::::::::::::::::
181 MetIleSerProArgMetValCysAlaGlyPheMetAspGlySerArgAs 197
483 CTCCTGACAGGATGATCTGGGGGGCCCTGATCTGACAGCGGTAACCT 532
   ::::::::::::::::::::
197 palAcysAsnGlyAspSerGlySerProLeuValCysArgGlyGlnValT 214
533 AGGGCTGTGTCTTGGGAAAGCCCGTGTGGCCACTTGGCGTGGCA 582
   ||| ::::::::::::::::::::
214 hrgIyLeuValSerTrpGlyGlnGly...CysAlaGlnProAsnTyrPro 229
583 GGTGTCTACCAACCTCTGCAAAATTCAGTAGTGATAGAGAAACCGT 632

```

```

||||| :::::::::::::::::::: |||
230 GYglValTyrValLeuIleAsnGlnTyrValLeuSerValAlaHisCysTrpT 246
633 C 633
246 u 246

```

seq_name: sp_vertibrate:092046

seq_documentation_block:
 ID 092046 PRELIMINARY; PRT; 249 AA.
 AC 092046; 013261;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, last annotation update)
 DE PREPROTEINASE PRECURSOR (EC 3.4.21.4) (ALPHA- AND BETA-TRYPSIN).
 OS Dissostichus mawsoni.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
 OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
 OC Perciformes; Notothenioidae; Nototheniidae; Dissostichus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PANCREAS;
 RA CHEN L., DEVRIES A.L., CHENG C.H.C.;
 RL submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 97268652.
 RA CHEN L., DEVRIES A.L., CHENG C.H.C.;
 RT "Evolution of antifreeze glycoprotein gene from a trypsinogen gene in
 RT Antarctic nototheniid fish."
 CC Proc. Natl. Acad. Sci. U.S.A. 94:3811-3816(1997).
 DR EMBL: U58945; AAB57732.1;
 DR EMBL: U58835; AAB57728.1;
 DR PFAM: PF00089; trypsin; 1.
 KW Signal; Hydrolase.
 FT SIGNAL 1 15 POTENTIAL.
 FT CHAIN 16 249 PROTRYPSIN.
 SQ SEQUENCE 249 AA; 27169 MW; 55C59EB4 CRC32;

alignment_scores:
 Quality: 447.00 Length: 216
 Ratio: 2.867 Gaps: 5
 Percent Similarity: 72.685 Percent Identity: 42.130

alignment_block:
 US-09-030-606-175 x 092046 ..

Align seg 1/1 to: 092046 from: 1 to: 249

```

1 GCGCAGCCCTGGCAGCGGCGACATGTCATGGAACGAAATGTTCTGCTC 50
   ::::::::::::::::::::
33 SerArgProTyrMetAlaSerLeuAsnTyrGlyTrpHis...PheCysG1 48
51 GGGCGCTCTGTGATCCGACAGTGGGTCTCTACGCCGACATGTTTCC 100
   ::::::::::::::::::::
48 YglValLeuIleAsnGlnTyrValLeuSerValAlaHisCysTrpT 65
101 AGACTCTCTACACC....ATCGGGCTGGGCGCTGCACAGTCTGAGGCC 144
   ||| ::::::::::::::::::::
65 YrAsnProTyrSerMetGlnValIleLeuGlnLysPheAsnLeuArgVal 81
145 GACCAAGAGCCGAGGAGCAGATGGTGGAGCGCCACTCTCCGTACGGCA 194
   ||| ::::::::::::::::::::
82 ...PheGlnGlyThrGlnLeuMetLysThrAspThrIleIleTrpH1 97
195 CCCAGATACACAGACTCTTGTCTGCTACGACCTCATGCTATCAAGT 244
   ||| ::::::::::::::::::::
97 sProSerTyrAspTyrGlnThrLeuAspPheAsnIleMetLeuIleYsL 114
245 TGAGAGATCGTGGTGGAGTCTGACACATCCGAGACATGATTCCT 294

```


DR EMBL; AB015206; BAA8895.1; -.
DR EMBL; Y18723; CAA77269.1; -.
DR PFAM; PF00089; trypsin. 1.
KW Protease; Serine protease.
SQ SEQUENCE 246 AA; 27496 MW; 90AC3E23 CRC32

```
alignment_scores:
  Quality: 441.50      Length: 212
  Ratio: 2.983         Gaps: 1
Percent Similarity: 69.811  Percent Identity: 38.679
```

alignment_block:
ns-00-030-606-175 -- 098301

Align seg 1/1 to: 088301 from: 1 to: 246

[illegible]

```

alignment_scores:
  Quality: 438.50      Length: 198
  Ratio: 2.943         Gaps: 3
  Percent Similarity: 75.253      Percent Identity: 40.909
alignment_block:
  US-09-030-606-175 x 093266 ..
  Align seg 1/1 to: 093266 from: 1 to: 242

```

```

alignment_scores:      quality: 438.50      length: 198
                       ratio:    2.943      gaps:    3
Percent Similarity:    75.253      Percent Identity: 40.905

alignment_block:
US-09-030-606-175 x G93266 ..

```

```
alignment_block:
US-09-030-606-175 x 093266 .
```

Align seg 1/1 to: 093266 from: 1 to: 242

[illegible]

```
seq_name: sp_vertebrate:O93266
seq_documentation_block:
ID: O93266 PRELIMINARY; PRT: 242 AA
```


193 GlngIaSPSeRgLyGlyProValIleCysAsnGlyGluLeuGlnGlyVa 209
 540 TGTGCTTTGGGAAAGCCCGCTGGCAACTGTGGCGTGGCAGGAGTCT 589
 209 IValSerTrpGlyTyrGly...CysAlaIleuArgGlyAsnProGlyVal 225
 590 ACACCAACCTCTGCAATTCAGTGTAGTATGAGAAAACCTC 633
 225 yTAlaLySValCysLeuPheAsnAspTrpLeuGluSerThrMet 239

seq_name: sp_vertebrate:042608

seq_documentation_block:
 ID 042608 PRELIMINARY; PRT; 247 AA.
 AC 042608;
 DT 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
 DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
 DE TRYPSINOGEN A3 PRECURSOR.
 GN TRYP3.
 OS Petromyzon marinus (Sea lamprey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Cephalaspidomorphi;
 OC Petromyzontiformes; Petromyzontidae; Petromyzon.
 RN [1]
 RA ROACH J.C.;
 RP SEQUENCE FROM N.A.
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RC TISSUE=ANTERIOR INTESTINE;
 RA ROACH J.C.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF011899; AAB6955.1;
 DR EMBL: AF011352; AAB65411.1;
 DR PRAW: PF00089; trypsin; 1.
 KW Signal.
 FT SIGNAL 1 15 POTENTIAL.
 FT CHAIN 16 247 TRYPSIN A3.
 SQ SEQUENCE 247 AA; 26295 MW; 6D71NC2E CRC32;

alignment_scores:
 Quality: 434.00 Length: 214
 Ratio: 2.800 Gaps: 5
 Percent Similarity: 72.430 Percent Identity: 39.720

alignment_block:
 US-09-030-606-175 x 042608 ..

Align seg 1/1 to: 042608 from: 1 to: 247

1 GCGCAGCCCTGGCAGCGGCGCACTGTGATGAAACGAATTGTCTCTC 50
 34 SerGlnProTrpGlnValSerLeuAsnIleGlyTyrHis...PheCysG1 49
 51 GGGCGCTCTGGTCATCCGATGGGTGCTGTCTGACCGCGACACTTTCC 100
 49 yGlySerLeuIleAsnSerGlnTrpValSerAlaIleHisCysTyrG 66
 101 AGAACTCC.....TACACCATGGGCTGGGCTGTGCACAGCTTGTAGGCC 144
 66 InThAlaSerArgIleSerValArgIleGlyGluHisAsnIlePheVal 82
 145 GACCAAGAGCCAGGAGCAGATGTTGAGAGCCAGCTCTCCGTACGGCA 194
 83 Asn...GluGlyThrGluGlnGlnIleGlnAlaSerLySAlaIleGlnH1 98
 195 CCCAGATAGACAGACACTTGTGCTGCTAAGACCTCATGCTCATCAGT 244
 98 SPTrGlnIlyrAsnSerTrpThrIleAspAsnSpleMetLeuIleLySL 115
 245 TGGAGAAATCCGTGTCCGAGTCTGACACCATCCGAGAGCATCAGATTGCT 294

115 euSerSerProAlaThrLeuAsnGlnIlyrAlaGlnAlaIleAlaLeuPro 131
 295 TCGAGACCCCTACCCCGCGGCACTTGTCTGCTTCTGCTGGGCT... 342
 132 SerSerCysValAsnThrGlyValMetCysThrIleSerGlyTrpGlyG1 148
 343 .CTGTGCGCAAGCAGCAAGAAATGCTACCGTCTGTGCTGCTGTAAGCTGT 391
 148 uThGlnThrSerValGlySerProAspValLeuMetCysValGlnAlaP 165
 392 CGGTGTGTCTGAGAGANGTCTGACATAAGCTCTATAGCCCGCTGTACAC 441
 165 rovalLeuSerAspThrSerCysArgAsnSerTyrProGlyAspIleThr 181
 442 CCCAGCATGTTCTGCGCGCGGAGGAGGCAACACGAAAGACTCTGAA 491
 182 AsnAsnMetIleCysLeuGlyTyrLeuGlnGlyGlyLySAspSerCysG1 198
 492 CGGTGACTCTGGGGGCGGCTGATCTGCACAGGCTACTTGCAGGCGCTTG 541
 198 nGlyAspSerIleGlyProValValCysAsnGlyGluLeuGlnGlyIleV 215
 542 TGTCTTTGGGAAAGCCCGCTGTGGCCAACTTGGCGTGCAGGTGTCTAC 591
 215 alSerTrpGlyArgGly...CysAlaLeuProAsnTyrProGlyValTyr 230
 592 ACACCACTCTGCAATTCAGTGTAGTATGAGAAAACCTC 633
 231 ThrLySValCysAsnTyrAsnAlaTrpIleAlaGlnIlyrIle 244

seq_name: sp_vertebrate:042158

seq_documentation_block:
 ID 042158 PRELIMINARY; PRT; 247 AA.
 AC 042158;
 DT 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
 DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
 DE TRYPSINOGEN A2 PRECURSOR.
 GN TRYP2.
 OS Petromyzon marinus (Sea lamprey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Cephalaspidomorphi;
 OC Petromyzontiformes; Petromyzontidae; Petromyzon.
 RN [1]
 RA ROACH J.C.;
 RP SEQUENCE FROM N.A.
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF011898; AAB6954.1;
 DR PRAW: PF00089; trypsin; 1.
 KW Signal.
 FT SIGNAL 1 15 POTENTIAL.
 FT CHAIN 16 247 TRYPSIN A2.
 SQ SEQUENCE 247 AA; 26309 MW; D049AA4C CRC32;

alignment_scores:
 Quality: 434.00 Length: 214
 Ratio: 2.800 Gaps: 5
 Percent Similarity: 72.430 Percent Identity: 39.720

alignment_block:
 US-09-030-606-175 x 042158 ..

Align seg 1/1 to: 042158 from: 1 to: 247

1 GCGCAGCCCTGGCAGCGGCGCACTGTGATGAAACGAATTGTCTCTC 50
 34 SerGlnProTrpGlnValSerLeuAsnIleGlyTyrHis...PheCysG1 49
 51 GGGCGCTCTGGTCATCCGATGGGTGCTGTCTGACCGCGACACTTTCC 100
 49 yGlySerLeuIleAsnSerGlnTrpValSerAlaIleHisCysTyrG 66

```

101 AGAAGTCC.....TACACCATCGGGCTGGGCGCTGCACAGTCTTGAGCC 144
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66 InthralSerArgIleSerValArgIleGlyIleHisAsnIlePheVal 82
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
145 GACCAAGAGCCAGGAGCCAGATGTTGGAGCCAGCTCCGTCAGGCA 194
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
83 Asn...GluGlyThrGluGlnGlnIleGlnIleAsnSerLysAlaIleGlnI 98
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
195 CCAGAGTACACAGACTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 244
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
98 sProGlnItyrAsnSerTrpThrIleAspAsnAspIleMetLeuIleYSL 115
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
245 TGAGCAGATCCGTGTCGAGCTGTGACACCATCCGAGCATCAGCATTCCT 294
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
115 euserSerProAlaThrIleuAsnGlnItyrAlaGlnAlaIleAlaLeuPro 131
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
295 TCAGCAGTCCCTACCGGGGAGACTTGTGCTGCTGCTGCTGCTGCTGCT 344
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
132 SerSerCysValAsnThrGlyValMetCysThrIleSerGlyTrpGlyG 148
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
345 GCTGGCGCAGCGGC...AGAATGCCCTACCGTGTGCTGCTGCTGCTGCT 391
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
148 uThrGlnIthSerIleGlySerProAspValLeuMetCysValGlnAlaP 165
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
392 CGGTGCTGTGTGAGGANGTCTGCAGTAGCTCTATGACCCGCTGTACAC 441
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
165 roValLeuSerAspThrSerCysArgAsnSerItyrProGlyAspIleThr 181
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
442 CCCAGCATGTTTCCGCCGCCGCGAGGCGCAGACAGACAGACTCTCTCAA 491
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
182 AsnAsnMetIleCysLeuGlyTyrLeuGlnGlyGlyLysAspSerCysG 198
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
492 CGGTGACTTGGGGGCGCCCTGATCTGCACAGCGGTACTTGCAGGCGCTTG 541
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
198 ngIyAspSerClyGlyProValValCysAsnGlyGluLeuGlnGlyIleV 215
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
542 TGCTTTTGGAAAAAGCCCGCTGGCGCAACTTGGCGTGCAGGTGTCTAC 591
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
215 aISerTrpGlyArgGly...CysAlaLeuProAsnItyrProGlyValItyr 230
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
592 ACCAAGCTCTGCAATTCAGTGAAGTATAGAGAAAACCGTC 633
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
231 ThrIyValCysAsnItyrAsnAlaTrpIleAlaGlnThrIle 244
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
seq_name: sp_rdent:Q9Z1R9
seq_documentation_block:
ID Q9Z1R9 PRELIMINARY; PRT; 246 AA.
AC Q9Z1R9;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE TRYSINOGEN 16
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RA ROMEN L., HOOD L.;
RT "Comparison between strains Balb/C and 129 in a region of the mouse T
cell receptor beta locus.";
RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF107342; AAC79093.1; -; 36625744 CRC32;
SQ SEQUENCE 246 AA; 26134 MW; 36625744 CRC32;

```

```

alignment_scores:
    Quality: 431.50      Length: 211
    Ratio: 2.839         Gaps: 4
    Percent Similarity: 72.038      Percent Identity: 39.336
alignment_block:

```

```

us-09-030-606-175 x Q9Z1R9 ..
Align seg 1/1 to: Q9Z1R9 from: 1 to: 246
7 CCCTGGCAGGCGGCACTGCTGATGGAAGAAAGAAATTTCTGCTGGGCGT 56
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
36 ProIyGlnValSerLeu...AsnSerGlyTyrHisPheCysGlyIlySe 51
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
57 CCGTGGGATCCGCACTGGGTGCTGCTGAGCCGCGACGCTTTTCCAAACT 106
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 rLeuIleAsnAspGlnTrpAlaValSerAlaAlaHisCysTyrItyrVal 68
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107 CTTACACATCGGGCTGGGCTGTCACAGTCTTGAGGCGCCAGCAAGCA 156
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
68 rglIleGlnValArgLeuGlyGlnHisAsnIleAsnVal...LeuGlyG 83
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
157 GGAAGCCAGATGTTGAGGCGCAGCTCTCCGTAAGCGCAGCAGAGTACA 206
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
84 AsnGlnGlnPheIleAspAlaAlaItyrIleItyrHisPheAsnPheAs 100
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
207 CAGACTCTTGTGCTGACACCATCCGAGCATCAGCATTCGTCGAGCATCG 256
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
100 nArgIyThrLeuAsnAsnAspIleMetLeuIleLysLeuSerProV 117
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
257 TGTCCAGTCTGACACCATCCGAGCATCAGCATTCGTCGAGCATCGCT 306
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
117 alThrIleuAsnAlaArgValAlaAlaThrValAlaLeuProSerSerCysAla 133
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
307 ACCGCGGGAGACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 350
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
134 ProAlaGlyThrGlnCysLeuIleSerGlyTrpGlyAsnThrLeuSerPh 150
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
351 GAAGCGAGAAATGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 400
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
150 eGlyValSerGlnProAspLeuGlnCysLeuAspAlaProLeuLeuP 167
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
401 CTGAGGANGTCTGCAGTAGCTCTATGACCCGCTGTACACCCACAGCANG 450
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
167 roGlnAlaAspCysGlnAlaSerTrpProGlyLysIleThrGlyAsnMet 183
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
451 TTTTCGCGCGCGGCGAGGCGCAGACAGCAGAAAGACTCTGCAACGGTACT 500
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
184 ValCysAlaGlyPheLeuGlnGlyGlyLysAspSerCysGlnGlySpSe 200
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
501 TGGGGGCGCCCTGATCTGCACAGGCTACTTGCAGGCGCTTGTGCTTGG 550
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
200 rGlyGlyProValValCysAsnGlyGluLeuGlnGlyIleValSerTrpG 217
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
551 GAAAGCCCGCTGTGCGCAACTTGGCGTGCAGGTGCTTACACCAACCTC 600
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
217 LtyrGly...CysAlaLeuProAspAsnProGlyValItyrThrIyVal 232
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
601 TGCAAATTCAGTGAAGTATAGAGAAAACCGTC 633
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
233 CysAsnItyrValAspTrpIleGlnAspThrIle 243
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:29:43 ; Search time 1809.22 Seconds
(without alignments)
1272.342 Million cell updates/sec

Title: US-09-030-606-175

Perfect score: 1167
Sequence: 1 GCGCAGCCCTGCGACGCGGC.....NTAGAGAGCGCAAAAAA 1167

Scoring table: IDENTITY_NUC
Searched: 2546578 segs, 986266752 residues

Database :
EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: em_est20: *
21: em_est21: *
22: em_est22: *
23: em_est23: *
24: em_est24: *
25: em_est25: *
26: em_est26: *
27: em_est27: *
28: em_est28: *
29: em_est29: *
30: em_est30: *
31: em_est31: *
32: em_est32: *
33: em_est33: *
34: em_est34: *
35: em_est35: *
36: em_est36: *
37: em_est37: *
38: em_est38: *
39: em_est39: *
40: em_est40: *
41: em_est41: *
42: em_est42: *
43: em_est43: *
44: em_est44: *
45: em_est45: *
46: em_est46: *
47: em_est47: *
48: em_est48: *
49: em_est49: *
50: em_est50: *
51: em_est51: *
52: em_est52: *
53: em_est53: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	555.6	47.6	777	48	AI557281 P12.1.15-
2	403.2	34.6	569	50	AI686688 t03g11.x
3	393	33.7	415	35	AA551449 n155605.s
4	348	29.8	404	34	AA503963 n139a01.s
5	286.6	24.6	576	34	AA533140 n146b06.s
6	221	18.9	722	48	AI557025 P12.1.10-
7	217.2	18.6	259	35	AA565161 n152c07.s
8	210.8	18.1	229	35	AA552779 n157a10.s
9	203	17.4	218	50	AI674521 w139g02.x
10	201.2	17.2	241	36	AA603977 n047a09.s
11	174	14.9	188	50	AI675815 w139g12.x
12	164	14.1	216	32	AA336074 EST4086
13	161.8	13.9	191	35	AA595489 n034n04.s
14	132	11.3	457	45	AI385433 m185b05.y
15	118.2	10.1	505	45	AI391329 mb71h07.y
16	116	9.9	496	39	AA684127 v142g07.r
17	114	9.8	496	39	AA684127 v142g07.r
18	112.6	9.6	507	35	AA583052 n080c09.s
19	106.4	9.1	517	31	AA293027 z154a12.r
20	100.2	8.6	590	33	AA411252 z135b03.r
21	99.2	8.5	759	45	AI326340 m161e08.x
22	99	8.5	321	31	AA302930 EST113097
23	98.4	8.4	686	46	AI415008 mb71h07.x
24	98	8.4	586	26	w13140 z155e11.s1
25	97.4	8.3	626	36	AA643312 n159d08.s
26	96	8.2	333	49	AI636241 t293c03.x
27	92.6	7.9	356	39	AA838788 a187f08.s
28	91.4	7.8	212	28	C16642 C16642 Clon
29	90.2	7.7	420	43	AI226226 w68f06.y
30	90	7.7	585	35	C23111 C23111 Japa
31	90	7.7	504	49	AI620091 t148a10.x
32	89.8	7.7	360	39	AA846771 e141f01.s
33	89.8	7.7	367	41	AI002101 c138d03.s
34	89.6	7.7	234	33	AA452459 z129g09.r
35	89	7.6	400	31	AA293231 z126g09.r
36	89	7.6	539	34	AA477689 z146a12.r
37	87.8	7.5	467	28	AA073833 m199h09.r
38	87.8	7.5	460	38	AA791893 v554h07.r
39	87.4	7.5	519	33	AA401397 z146b01.s
40	87.4	7.5	583	40	AA921373 a156d12.s
41	87.4	7.5	493	45	AI324874 m185b05.x
42	87.2	7.5	541	39	AA844955 a161d09.s
43	87.2	7.5	324	43	AI177474 EST221106
44	87.2	7.5	599	43	AI237604 EST234166
45	87.2	7.5	457	46	AA998593 UT-R-CO-1

ALIGNMENTS

RESULT 1
LOCUS AI557281 777 bp mRNA
DEFINITION PT2.1.15.G12.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION AI557281
NID 94489644
VERSION AI557281.1 GI:4489644
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS 1 (bases 1 to 777)
Huang, G.M., Ng, W., Farakas, J., Chen, L., Liang, H.A., Gordon, D., Jun Yu, J., and Hood, L.
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138767.

Contact: Guyang Mathew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huangm@yahoo.com.

FEATURES
source
1..777
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="960H11: 6; 6p21.31-6p22.1; 21q"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional cDNA library was constructed using Lambda ZP II kit (stratagene). mRNA was extracted from a frozen prostate tumor tissue (Mayo Clinics)."

BASE COUNT 162 a 205 c 227 g 157 t 26 others
ORIGIN

Query Match 47.6%; Score 555.6; DB 48; Length 777;
Best Local Similarity 91.8%; Pred. No. 1.3e-128;
Matches 637; Conservative 0; Mismatches 46; Indels 11; Gaps 5;

2 CGCAGCCCTGGCAGCGGCGCAGTGTGATGAAAAAGATTTGCTCGCGGCGTCTCG 61
|||||
21 CGCAGCCCTGGCAGCGGCGCAGTGTGATGAAAAAGATTTGCTCGCGGCGTCTCG 80
|||||
62 TGCATCCGACAGTGGGTCTGTCAAGCCGACAGTGTTCAGAACTCTTACACCATCGGGC 121
|||||
81 TGCATCCGACAGTGGGTCTGTCAAGCCGACAGTGTTCAGAACTCTTACACCATCGGGC 140
|||||
122 TGGGCGCTGGCAGCACTGTGAGGCGGACCAAGCGAGGAGCCAGATGTGAGGCCAGCC 181
|||||
141 TGGGCGCTGGCAGCACTGTGAGGCGGACCAAGCGAGGAGCCAGATGTGAGGCCAGCC 200
|||||
182 TCTCCGTACGCGACCCAGAGTATACAGAGCTTGTGCTGCTAGAGCCTATGCTATCA 241
|||||
201 TCTCCGTACGCGACCCAGAGTATACAGAGCTTGTGCTGCTAGAGCCTATGCTATCA 260
|||||
242 AGTTGAGCAATCCGTGTCGAGTCTGACACCATCCGAGACATCAGATTGCTTGCAGT 301
|||||
261 AGTTGAGCAATCCGTGTCGAGTCTGACACCATCCGAGACATCAGATTGCTTGCAGT 320
|||||
302 GGCCTACCGGGGGAACTCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 361
|||||
321 GGCCTACCGGGGGAACTCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 380
|||||
362 TGCCTACCGGTCTGACACTGCTGTAACGTGCGGTGCTGCTGAGAGAGTCTGAGTAAGC 421
|||||
381 TGCTACCGGTCTGACACTGCTGTAACGTGCGGTGCTGCTGAGAGAGTCTGAGTAAGC 439
|||||
422 TCTATGACCGGTCTGACACTGCTGTAACGTGCGGTGCTGCTGAGAGAGTCTGAGTAAGC 481
|||||
440 TCTATGACCGGTCTGACACTGCTGTAACGTGCGGTGCTGCTGAGAGAGTCTGAGTAAGC 499
|||||
482 ACTCCGTGACAGGTGACTGCGGGGGGCGCTGATCTCAACGGGTACTTGAGAGGCGCTTG 541
|||||
500 ACTCCGTGACAGGTGACTGCGGGGGGCGCTGATCTCAACGGGTACTTGAGAGGCGCTTG 558
|||||

QY 542 TGTCTTGGAAAAAGCCCGTGGGCCAATGCGCGTCCAGGTGCTACACCAACCTCT 601
|||||
DB 559 TGTCTTGGAAAAAGCCCGTGGGCCAATGCGCGTCCAGGTGCTACACCAACCTCT 611
|||||
QY 602 GCAATTCACGTAGTGGATAGAGAAAAAGCCGTCAGGCCAGTTAATCTGGGAGTGGAA 661
|||||
DB 612 CTGAATTAATCTGAGTGGATAGAGAAAAAGCCGTCAGGCCAGTTAATCTGGGAGTGGAA 669
|||||
QY 662 CCATGAATTAATGCCCCCAATTAATCAATCCGCGGA 695
|||||
DB 670 NCCATTAATTAATGNCACCAAAATTAATCTGGGGAA 703
|||||

RESULT 2
A1686689/c
LOCUS
DEFINITION
A1686689 569 bp mRNA EST 27-MAY-1999
tuz3g11.x1 NCI-CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2253092 3'
similar to TR:Q92046 Q92046 PREPROTRYSIN PRECURSOR ;, mRNA
sequence.
ACCESSION A1686689
NID 94897983
VERSION A1686689.1 GI:4897983
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 569)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
TITLE Unpublished (1997)
JOURNAL On Jun 5, 1998 this sequence version replaced gi:3189584.
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -400P from Glibco
High quality sequence stop: 444.
Location/Qualifiers
1..569
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2253092"
/clone_lib="NCI-CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pTRF3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 125 a 150 c 170 g 124 t
ORIGIN

Query Match 34.6%; Score 403.2; DB 50; Length 569;
Best Local Similarity 98.8%; Pred. No. 1.1e-90;
Matches 405; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 151 GAGCAGGAGGAGCCAGATGTGTGAGAGCCAGCCTCTCCGTACGGACACCAGAGTACAAAGA 210
Db 569 GAGCAGGAGGAGCCAGATGTGTGAGAGCCAGCCTCTCCGTACGGACACCAGAGTACAAAGA 510
QY 211 CTCCTGCTCGCTAACGACCTCATGCTCATCAATTGGAGAAATCCGTGTCCGAGTCTAC 270
Db 509 CCGTGTCTCGCTAACGACCTCATGCTCATCAATTGGAGAAATCCGTGTCCGAGTCTAC 450
QY 271 ACCATCCGGAGACATCAGCATTTGCTTGGCAAGTCCCTACCGCGGGGAACTCTTGCTGCTN 330
Db 449 ACCATCCGGAGACATCAGCATTTGCTTGGCAAGTCCCTACCGCGGGGAACTCTTGCTGCTT 390
QY 331 TCTAGCTGGGGTGTGCTGTGGGGAACGGAGAAATCCATACCGTCTGTGACTGTGGGTGAAGTG 390
Db 389 TCTAGCTGGGGTGTGCTGTGGGGAACGGAGAAATCCATACCGTCTGTGACTGTGAAGTG 330
QY 391 TCGGTGGTGTCTGTAGGANGTCTGCAGTAAAGCTGTATGACCCGCTGTACCAACCCAGCATG 450
Db 329 TCGGTGGTGTCTGTAGGANGTCTGCAGTAAAGCTGTATGACCCGCTGTACCAACCCAGCATG 270
QY 451 TTCTGCGCCGCGGAGGGGCAAGACCCAGAAAGAGACTCTGCAACGCTGACTCTGGGGGGCCC 510
Db 269 TTTTGGCCCGCGGAGGGGCAAGACCCAGAAAGAGACTCTGCAACGCTGACTCTGGGGGGCCC 210
QY 511 CTGATCTGCAACGGGTACTTGCAAGGGCGTTGTCTTTTGGAAAAAGCCC 560
Db 209 CTGATCTGCAACGGGTACTTGCAAGGGCGTTGTCTTTTGGAAAAAGCCC 160

RESULT	3
LOCUS	AA551449
DEFINITION	AA551449 415 bp MRNA EST 05-SEP-1997 P155805.s1 NCI CGAP Pr9 Homo sapiens CDNA clone IMAGE:996416 similar to SW:IKK_MOUSE P15946 GLIANDULAR KALLIKREIN KII PRECURSOR ,, mRNA sequence.
ACCESSION	AA551449
NID	G2321701
VERSION	AA551449.1 GI:2321701
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homalidae; Homo. 1 (bases 1 to 415)
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	On May 9, 1995 this sequence version replaced gi:802243.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuagui,
 M.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: David B. Kitzman, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/URL at:
www-bio.liml.gov/dbfp/image/image.html
 Insert Length: 640 Std Error: 0.00
 Seq primer: -40mls fwd. ET from Amersham
 High quality sequence stop: 412.
 Location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:966416"
/clone_1db="NCI_CGAP_P19"
/sex="male"

```

```

/tissue_type="normal prostatic epithelial cells"
/lab_host="DH108"
/note="Organ: prostate; Vector: PAMPI10: mRNA made from
normal prostatic epithelial cells. cDNA made by oligo-dT
priming. Non-directionally cloned. Size-selected on
agarose gel, average insert size 600 bp. Library made by
D. Krlleman, NIH."
BASE COUNT      78 a      125 g      83 t
ORIGIN          125 c

```

Query Match	Similarity	Score	DB	Length
Best: Local	98.1%	Pred. No. 3.3e-88;		
Matches: 407;	Conservative	0;	Mismatches	7;
			Indels	1;
			Gaps	1;

QY	137	TTGAGCGCCGACCAAGAGCCAGAGGAGCCAGAGAGTGtGAGAGCCAGCCTCTCCGTACGGCAC	196
Db	2	TCGAGCGCCGACCAAGAGCCAGAGGAGCCAGAGGAGCCAGCCTCTCCGTACGGCAC	61
QY	197	CAGAGTACAAAGACTCTGCTCGCTAACGACCCTCATGCTCTATCAAGTTGGACGAATCCG	256
Db	62	CAGAGTACAAAGAGCCTCTGCTCTAACGACCCTCATGCTCTATCAAGTTGGACGAATCCG	121
QY	257	TGTCCGAGTCTGAACACATCCGAGACATCACATTTGCTTCAGAGCCCTACCGCGGGGA	316
Db	122	TGTCCGAGTCTGAACACATCCGAGACATCACATTTGCTTCAGAGCCCTACCGCGGGGA	181
QY	317	ACTCTGCGCTGCTGTCGTGCGTGGGGTCTGCTGGCGAAGCGCAATGCTTACCGTGTGC	376
Db	182	ACTCTGCGCTGCTTCTGCGTGGGGTCTGCTGGCGAAGCGCAATGCTTACCGTGTGC	240
QY	377	ACTCGTGAACGTTCGGTGGTGTCTGTAGAGANGTCTGAGTAAGCTATATACCCGGCTGT	436
Db	241	AGTGGTGAACGTTCGGTGGTGTCTGTAGAGANGTCTGAGTAAGCTATATACCCGGCTGT	300
QY	437	ACCAACCCAGCATGTTCTGCGCGCGCGGAGGGCAAGACAGAGAGACTCTGCAACGGTG	496
Db	301	ACCAACCCAGCATGTTCTGCGCGCGCGGAGGGCAAGACAGAGAGACTCTGCAACGGTG	360
QY	497	ACTGTGGGGGCCCTGATCTGCAAGGGGTACTTTCAGAGGCTGTGTCTTTCGG	551
Db	361	ACTGTGGGGGCCCTGATCTGCAAGGGGTACTTTCAGAGGCTGTGTCTTTCGG	415

RESULT 4
LOCUS AA503963/c
DEFINITION
AA503963 404 bp mRNA EST 20-AUG-1997
nm39a01.s1 NCI_CGAP_P5 Homo sapiens CDNA clone IMAGE:954656
similar to SW:K1KA_MOUSE P15946 GLIADULAR KALLIKREIN K11 PRECURSOR
;contins MSRL.b2 MSRL repetitive element ;, mRNA sequence.
AA503963
NID g2238830
VERSION AA503963.1 GI:2238930
KEYWORDS
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Euthetia; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 404)
REFERENCE
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
COMMENT
On May 8, 1995 this sequence version replaced gi:801235.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuagui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Kitzman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/dbp/image/image.html

Insert Length: 1244 Std Error: 0.00
Seq primer: -40m13 fwd. ET from AmerSham
High quality sequence stop: 144.
Location/Qualifiers

FEATURES

source

1. 404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:954696"
/clone_1ib="NCI-CGAP_Pr5"
/sex="male"
/tissue_type="prostate"
/lab_host="DH10B"
/note="Vector: PAMP10; mRNA made from normal prostatic epithelial cells, CDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp."
BASE COUNT 64 a 91 c 159 g 90 t
ORIGIN

Query Match 29.8%; Score 348; DB 34; Length 404;
Best Local Similarity 93.5%; Pred. No. 5.3e-77;
Matches 371; Conservative 0; Mismatches 25; Indels 1; Gaps 1;

QY 502 GGGGGGGGGGCTGATCTGCACAGGGGCTTGTGTTCGGAAGAGCCCG 561
|||||
DB 404 GGGGGGGGGGCTGATCTGCACAGGGGCTTGTGTTCGGAAGAGCCCG 345
QY 562 TGTGGCAACTTGGCGTGCAGAGTGTCTACACCAACTCTGCAATTC-CTGAGTGGAT 620
344 TGTGGCAACTTGGCGTGCAGAGTGTCTACACCAACTCTGCAATTCAGTGGAT 285
QY 621 AGAGAAAACCGTCCAGNCAGTTACTCTGGGAGCTGGGAACCATGAATGACCCCA 680
DB 284 AGAGAAAACCGTCCAGNCAGTTACTCTGGGAGCTGGGAACCATGAATGACCCCA 225
QY 681 AATACATCTCGCGGAAGAAATTCAGAAATCTGTTCGCCAGCCCTCTCTCCTAGGCC 740
224 AATACATCTCGCGGAAGAAATTCAGAAATCTGTTCGCCAGCCCTCTCTCCTAGGCC 165
DB 741 AGAGTCCAGAGCCCGGAGTCCAGAGCCCGGAGTCCAGAGTCCAGAGTCCAG 800
DB 164 AGAGTCCAGAGCCCGGAGTCCAGAGCCCGGAGTCCAGAGTCCAGAGTCCAG 105
QY 801 CTCCTCAGACCCAGAGTCCAGAGCCCGGAGTCCAGAGTCCAGAGTCCAGAGTCC 860
DB 104 CTCCTCAGACCCAGAGTCCAGAGCCCGGAGTCCAGAGTCCAGAGTCCAGAGTCC 45
QY 861 GGGGCTCTCTCCNTCAGACGAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCC 897
DB 44 GGGGCTCTCTCCNTCAGACGAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCC 8

RESULT 5

AA533140

LOCUS AA533140 576 bp mRNA EST 21-AUG-1997
DEFINITION nj46106.s1 NCI-CGAP_Pr5 Homo sapiens cDNA clone IMAGE:995579
similar to TR:G198491 G198491 KALLIKREIN; contains MSRI.tl MSRI
repetitive element; mRNA sequence.

ACCESSION

AA533140
NID 92277236
VERSION AA533140.1 GI:2277236
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
1 Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
On Sep 12, 1996 this sequence version replaced gi:1406821.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Lihnan, M.D., Rodrigo F. Chuagui,
M.D., Michael R. Emmerit-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/dbp/image/image.html

Insert Length: 885 Std Error: 0.00
Seq primer: -40m13 fwd. ET from AmerSham
High quality sequence stop: 256.
Location/Qualifiers

FEATURES

source

1. 576
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:95579"
/clone_1ib="NCI-CGAP_Pr9"
/sex="male"
/tissue_type="normal prostatic epithelial cells"
/lab_host="DH10B"
/note="Organ: prostate; Vector: PAMP10; mRNA made from normal prostatic epithelial cells, CDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Library made by D. Krizman, NIH."
BASE COUNT 169 a 156 c 139 g 102 t 10 others
ORIGIN

Query Match 24.6%; Score 286.6; DB 34; Length 576;
Best Local Similarity 94.7%; Pred. No. 1.2e-61;
Matches 306; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 472 GACCAAGAGACTCTGCAACGGTGA-CTGGGGGGGGGCTGATCTGCACGGGTACTT 530
|||||
DB 42 GACCAAGAGACTCTGCAACGGTGA-CTGGGGGGGGGCTGATCTGCACGGGTACTT 101
QY 531 GCGAGGCTTGTCTTGTGGAAAAGCCCGTGTGCGCAACTTGGCTGCGAGGTCTTA 590
DB 102 GCGAGGCTTGTCTTGTGGAAAAGCCCGTGTGCGCAACTTGGCTGCGAGGTCTTA 161
QY 591 CACCAACTCTGCAATTCAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGT 650
DB 162 CACCAACTCTGCAATTCAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGT 221
QY 651 GGGAGTGGAAACCATGAATTTGACCCCAATATATCTCTGGGAAGAATTCAGGAATA 710
DB 222 GGGAGTGGAAACCATGAATTTGACCCCAATATATCTCTGGGAAGAATTCAGGAATA 281
QY 711 TCTGTTCCAGAGCCCTCTCTCTCAGAGCCAGAGTTCAGAGCCCGGAGTCTCTCTCT 770
DB 282 TCTGTTCCAGAGCCCTCTCTCTCAGAGCCAGAGTTCAGAGTTCAGAGTTCAGTCTG 341
QY 771 CAACCAAGGTTACAGTCCCA 793
DB 342 TCAACCAAGGTTACAGTTCCA 364

RESULT 6

A1557025

LOCUS A1557025 722 bp mRNA EST 23-MAR-1999
DEFINITION FT2.1_10_F05.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION A1557025
NID 94489388

VERSION	AA565161.1	GI:2336800
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	

QY	915	ACCCAGGGGACAGAGCCGCCAACCCCTC	NTCMTTCAGATGAGAGGTCCCAAGCCCC	AAC	974
Db	245	AACCAAGGTACAGATCCCAAGCCCTCCT	CCTTCAGACCCAGAGTCCCAAGCCCC	AAC	186
QY	975	CCCTGCTTCCCAAGACCCAGAGGTMCAG	GTCCAGGCCCTCCTCCTCAGACCCAGCG	GT	1034
Db	185	CCCTCCTTCCCAAGACCCAGAGGTCCAG	GTACAGGCCCTCCTCCTCAGACCCAGCG	GT	126
QY	1035	CCATGCGACCTAGANTNCCCGTACAGAG	TCCCGCTGTGGCAGNTGACCCCAACT		109
Db	125	CCAAATGCACTGACACTCTCCGTACATG	TGCCCCCTGTGGGACGTTGACCCCACT		66
QY	1095	TACCAAGTTGGTTTTTCATTTTTTTGT	TCCCTTCCCTAGATCCAGAAATTAAG	TTAAAG	1154
Db	65	TACCAAGTTGGTTTTTCATTTTTTTGT	TCCCTTCCCTAGATCCAGAAATTAAG	TTAAAG	6
QY	1155	AACGC	1159		
Db	5	AAGCG	1		

RESULT	8
AA552779/c	
LOCUS	AA552779
DEFINITION	229 bp mRNA EST 08-SEP-1997 nk57a10.s1 NCI-CCAP Pr7 Homo sapiens cDNA IMAGE:1017594 similar to contains MSRI.b3 MSRI repetitive element ;, mRNA sequence.

ACCESSION AA552779
 NID 92323033
 VERSION AA552779.1 GI:2323033
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 229)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL On Jan 25, 1995 this sequence version replaced gi:637760.
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuagui,
 M.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: David B. Krizman, Ph.D.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbrrp/image/image.html
 Insert Length: 419 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 215.
 Location/Qualifiers
 1..229
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1017594"
 /clone_1ib="NCI_CGAP_pr7"
 /sex="male"
 /tissue_type="low-grade prostatic neoplasia"
 /note="Organ: prostate; Vector: pAMP10; mRNA made from
 prostate intraepithelial neoplasia (high-grade), cDNA
 made by oligo-dT priming. Non-directionally cloned.
 Size-selected on agarose gel, average insert size 600 bp."
 BASE COUNT 49 a 38 c 94 g 48 t
 ORIGIN
 Query Match 18.1%; Score 210.8; DB 35; Length 229;
 Best Local Similarity 93.9%; Pred. No. 6.5e-43;
 Matches 215; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
 QY 931 CCCCAACCCCTGTCCTGAGTCCAGAGTCCAGAGCCCAACCCCTGTCCTCCACAG 990
 DB 229 CCCCAACCCCTGTCCTGAGTCCAGAGTCCAGAGCCCAACCCCTGTCCTCCACAG 170
 QY 991 CCAGAGTCCAGTCCAGAGCCCTGCTCCCTCAGAGCCAGGCTCCATGCCACTAGAN 1050
 DB 169 CCAGAGTCCAGTCCAGAGCCCTGCTCCCTCAGAGCCAGGCTCCATGCCACTAGAC 110
 QY 1051 TATCCCTGACAGAGTCCCTGCTGAGCAGTGTGACCACTTACAGTGTGTTTC 1110
 DB 109 TATCCCTGACAGAGTCCCTGCTGAGCAGTGTGACCACTTACAGTGTGTTTC 50
 QY 1111 ATTTTGTCCCTTCCCTCAGATCCAGAAATTAAGTNTAAGAGAGCG 1159
 DB 49 ATTTTGTCCCTTCCCTCAGATCCAGAAATTAAGTNTAAGAGAGCG 1
 RESULT 9
 LOCUS A1674521 218 bp mRNA EST 19-MAY-1999
 DEFINITION wg39g02.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2321042 3'

similar to contains MSRL.t2 MSRL repetitive element ;, mRNA
 sequence.
 ACCESSION A1674521
 NID 94875001
 VERSION A1674521.1 GI:4875001
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 218)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL On Mar 16, 1998 this sequence version replaced gi:2961915.
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbrrp/image/image.html
 Seq primer: -40UP from Gibco.
 Location/Qualifiers
 1..218
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2321042"
 /clone_1ib="NCI_CGAP_Pr28"
 /sex="male"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
 with a modified polylinker; Plasmid DNA from the
 normalized library NCI_CGAP_Pr22 was prepared, and ss
 circles were made in vitro. Following HAP purification,
 this DNA was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from a pool
 of 5,000 clones made from the same library (clones
 985608-986759, 1101192-1101959, and 1217928-1220615)."
 Subtraction by Bento Soares and M. Fatima Bonaldo.
 BASE COUNT 48 a 36 c 85 g 49 t
 ORIGIN
 Query Match 17.4%; Score 203; DB 50; Length 218;
 Best Local Similarity 94.9%; Pred. No. 5.6e-41;
 Matches 206; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 QY 944 TCCNTCAGAGTCCAGAGTCCAGAGCCCAACCCCTGTCCTCCAGAGCCAGGTCACG 1003
 DB 217 TCCCTCAGAGTCCAGAGTCCAGAGCCCAACCCCTGTCCTCCAGAGCCAGGTCACG 158
 QY 1004 TCCAGAGCCCTGCTCCCTCAGAGCCAGCGGTCCATGACCTAGANTTCCCTGTAC 1063
 DB 157 TCCAGAGCCCTGCTCCCTCAGAGCCAGCGGTCCATGACCTAGANTTCCCTGTAC 98
 QY 1064 AGTGGCCCTGTTGGCAGNTGACCAACCTTACAGAGTGTGTTTCATTTTGTCCCT 1123
 DB 97 AGTGGCCCTGTTGGCAGNTGACCAACCTTACAGAGTGTGTTTCATTTTGTCCCT 38
 QY 1124 TTCCCTAGATCCAGAAATTAAGTNTAAGAGAGCGC 1160
 DB 37 TTCCCAAGATCCAGAAATTAAGTNTAAGAGAGAGCGC 1

RESULT	10
AA603977/c	
LOCUS	
DEFINITION	AA603977 241 bp mRNA EST 28-OCT-1997 nc67a99.s1 NCI-CGAP_Pt23 Homo sapiens CDNA clone IMAGE:1103800 3

ACCESSION	AA603977	
NID	g2444547	
VERSION	AA603977.1	GI:2444547
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE 1 (bases 1 to 241)
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1395285.

Contact: Robert Strausberg, Ph.D.
Tel.: (201) 406-1550

Tel.: (301) 496-1330
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Stratagene, Inc.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
[www-bio.linn.gov/DBP/Image/Image.html](http://www.bio.linn.gov/DBP/Image/Image.html)

```

Insert length: 2084      Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 231.
      Location/Qualifiers
source          1..241

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1103800"
/clone_1bp="NCI_CGAP_P123"
/sex="male, pooled"
/tissue_type="prostate tumor"
/lab_host="SOLR (kanamycin resistant)"
/ncore="Organ: prostate; Vector: Bluescript SK-; Site: 1;
EcoRI: Site 2; XhoI: Cloned unidirectionally. Primer:
Oligo dT. Pooled prostate tumors. 5' adaptor sequence: 5'
GAATTCGCACGAC 3' 3' adaptor sequence: 5'
CTCAGCTTTTCTTTTCTTTT 3' Average insert size: 1.2 kb."

```

Query Match	Score	DB	Length
17.28;	201.2;	36;	241;
06.28;	160.40;		

Matches 203; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 957 GAGGTCCAAGCCCCCAACCCCTGTTCCCCAGACCCAGAGGTTCAAGGTCCAGGCCCTTC 101a
|||||
Db 241 GAGGTCCAAGCCCCCAACCCCTCTTCCCCAGACCCAGAGGTTCAAGGTCCAGGCCCTTC 182

1017 TCCCTCAGACCCAGCGGTCCATGCGCACCTAGANTTCCCTGTACACAGTGCCCCCTTGT 1076

Db 181 TCCCTCAGACCCAGCGGTCCAATGCCACCTAGACTCTCCCTGTAAACAGTGCCCCCTTGT 122

1077 GGCANGTGCACCAACCTTACCAGTTGGTTTTCATTTTGTCCCTTCCCCCTAGATCC 1136

Db 121 GGCACGTTGACCCAACCTTACCAGTTGGTTTTTCATTTTTTGTCCCTTTCCCTAGATCC 62

113 / AGAAATAAAGTNTTAAAGAGAGCGCAAAAAA 116 /

Db 61 AGAATAAGTCTAGAGAGCGCAAAAAA 31

RESULT 11
AT675815/c

ACCESSION	Al675815	mrna sequence.
NID	94876295	
VERSION	Al675815.1	GI:4876295
KEYWORDS	EST.	
SOURCE	human.	

REFERENCE
AUTHORS
TITLE
JOURNAL

Eutheria: Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 188)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP)
Tumor Gene Index
Unpublished (1997)

[illegible]

Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/ILNLT at:
www-bio.litl.gov/bdnp/image/image.html

FEATURES	Location/Qualifiers
source	1. .188

```

BASE COUNT
ORIGIN
43 a
31 c
53 t
Subtraction by Bento Soares and M. Fatima Bonaldo.
985608-986759, 1101192-1101959, and 1217928-1220615.
5,000 clones made from the same library (cloneds
reaction. The driver was PCR-amplified cDNAs from a pool
this DNA was used as tracer in a subtractive hybridization
normalized library NCI_CGAP_Pr22 was prepared, and ss
circulars were made in vitro. Following HAP purification,
/note=Organ: prostate; Vector: pT73d-Pac (Pharmacia)
/ldb_host="DH10B"
/ldb_stage="adult"
/sex="male"
/clone_1lb="NCI_CGAP_Pr28"
/clone_image="2313599"
/db_xref="taxon:9606"
/organism="Homo sapiens"

```

Query Match	14.98; Score 174; DB 50; Length 188;
-------------	--------------------------------------

Matches 174; Conservative 0; Mismatches 5; Indels 0; Gaps 0

989 ACCCAGAGGTNCAGGTCCCGCCCTCTCCCTCAGACCCAGCGGTCCAATGCCACCTAG 1048

Db 188 ACCCAGAGGTCAGGTCGCCAGCCCTCCTCCCTCAGACCCAGGGTCCAATGCCACCTAG 129

1049 ANTNTCCCTGTACACAGTGCCCCCTTGTGGCANGTGCACCAACCTTACCAGTTGGTTT 1108

Db 128 ACCTCCCTGTACACAGTGCCTTGTGGCAGTTGACCCAACCTTACCAGTTGTTT 6

1109 TCAATTTTGTCCCTTTCCTCAGATCCAGAAATAAGGTNTAAGAGAGCCGCAAAAAA 11

DD 00 1C A11111G1C111C1C1A G A1C C A G A A A1 A A A G111 A A G A G A1B C C A A A A A A A A 10

RESULT 12
AA336074 216 bp mRNA EST 21-APR-1997
LOCUS ESR40886 Endometrial tumor Homo sapiens CDNA 5' end similar to
DEFINITION similar to kallikrein family, mRNA sequence.
ACCESSION AA336074
NID 91988560
VERSION AA336074.1 GI:1988560
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 216)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gockayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glocke,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaro,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,R.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferris,A., Fischer,C., Hastings,G.A.,
He,M.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,O.L., Kunsch,C., Hungjun,J., Li,H., Meisner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Rubin,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
JOURNAL Initial assessment of human gene diversity and expression patterns
MEDLINE based upon 83 million nucleotides of cDNA sequence
COMMENT 96026280
On Apr 14, 1993 this sequence version replaced gi:692773.

CONTACT: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M3 Reverse.
Location/Qualifiers
FEATURES
source 1..216
/organism="Homo sapiens"
/db_xref="ATCC (Inhost):138014"
/db_xref="taxon:9606"
/clone_lib="Endometrial tumor"
/sex="female"
/dev_stage="adult"
/note="Organ: endometrium; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 45 a 65 c 63 g 38 t 5 others
ORIGIN
Query Match 14.1%; Score 164; DB 32; Length 216;
Best Local Similarity 97.0%; Pred. No. 2.9e-31;
Matches 164; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CGAGCCCTGCGAGCGGCGGCTGTCATGGAAGAATGTCGTCGGGGCTCTGG 61
|||||
DB 48 CGAGCCCTGCGAGCGGCGGCTGTCATGGAAGAATGTCGTCGGGGCTCTGG 107
|||||
QY 62 TGCATCCGAGTGGGTGCTGTCTGACCGGACACATGTTCCGAACTCTACCATCGGGC 121
|||||

DB 108 TGCATCCGAGTGGGTGCTGTCTGACCGGACACATGTTCCGAACTCTACCATCGGGC 167
QY 122 TGGGCTGCGACGCTCTGAGCGCCGACCAAGCCGAGGAGCGAGATGGT 170
|||||
DB 168 TGGGCTGCGACGCTCTGAGCGCCGACCAAGCCGAGGAGCGAGATGGT 216
|||||

RESULT 13
AA595489 191 bp mRNA EST 18-SEP-1997
LOCUS n034h04.s1 NCI_CGAP_P123 Homo sapiens CDNA clone IMAGE:1102615 3',
DEFINITION mRNA sequence.
ACCESSION AA595489
NID 92410839
VERSION AA595489.1 GI:2410839
KEYWORDS EST
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 191)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Apr 14, 1993 this sequence version replaced gi:692625.

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNI at:
www.bio.lnl.gov/db/rrp/image/image.html
Seq primer: -40m13 fwd. ET from Amersham.
Location/Qualifiers
FEATURES
source 1..191
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1102615"
/clone_lib="NCI_CGAP_P123"
/sex="male, pooled"
/tissue_type="prostate tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: prostate; Vector: Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pooled prostate tumors. 5' adaptor sequence: 5'
GAATTCGGCAGAG 3' 3' adaptor sequence: 5'
CTCGAGTCTTTTCTTTTCTTTT 3' Average insert size: 1.2 kb."
BASE COUNT 43 a 28 c 57 g 63 t
ORIGIN
Query Match 13.9%; Score 161.8; DB 35; Length 191;
Best Local Similarity 95.9%; Pred. No. 9.8e-31;
Matches 163; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 998 TTAGGTCGCGAGCCCTCTCTCTCTGACGCGGCGGCTGTCATGCGAGANTTCCCT 1057
|||||
DB 191 TCGAGTACGACCCCTCTCTCTCTGACGCGGCGGCTGTCATGCGAGANTTCCCT 132
|||||

QY 1058 GTACACAGTGGCCCTTGGGANGTGCACCACTTACAGTGTGTTTCATTTT 1117
|||||
DB 131 GTACACAGTGGCCCTTGGGANGTGCACCACTTACAGTGTGTTTCATTTT 72
|||||

QY 1118 GTCCCTTCCCTAGATCCAGAAATTAAGTNTAAGAGAGCGCAAAAAA 1167
|||||
DB 71 GTCCCTTCCCTAGATCCAGAAATTAAGTNTAAGAGAGCGCAAAAAA 22
|||||


```
/clone_11b="Soares_fetal_heart_NBH19W"  
/sex="unknown"  
/dev_stage="19 weeks"  
/lab_host="DH10B (ampicillin resistant)"  
/note="Organ: heart; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTACCAATCTGAGGAGGCGCGCATCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by  
M.Fatima Bonaldo. This library was constructed from the  
same fetus as the fetal lung library, Soares fetal lung  
NBH19W."  
BASE COUNT      116 a      150 c      128 g      104 t  
ORIGIN
```

```
Query Match      10.1%; Score 118.2; DB 26; Length 498;  
Best Local Similarity 61.0%; Pred. No. 1e-19;  
Matches 227; Conservative 0; Mismatches 136; Indels 9; Gaps 2;
```

```
QY 275 TCCGAGCATCAGCATGCTTCGACGTGCCCTACCGGGGAACTCTTGCCCTGTCG 334  
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 25 TCAGACCCATCAACGCTCTCTCATTTGCTGCTGGGACAAAGTGTGTGTCG 84  
QY 335 GCTGGGGTCTGCTGGGAGGAGCAGATG-----CCTACCGTCTGCACTGGCTGACG 388  
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 85 GCTGGGGAGAACCAAGAGCCCAAGTGACCTTCCTAAGGTCTCCAGTCTTGATA 144  
QY 389 TGTGCGTGTGTGTGAGGANGTCTGCAAGTCTATGACCCGCTGACCCAGCA 448  
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 145 TCAGCCTGCTAACTCAGAAAGGTGAGAGATCTTACCGAGACAGATGATGACACCA 204  
QY 449 TGTTCGCGCGCGGAGGAGGAGACCAAGAGACTCTGCAAGGCTGACTGTGGGGGC 508  
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 205 TGTTCGCGCGCGGTGACA---AAGCAGGTGAGAGACTCTCCAGGGTGTGAGGGGC 261  
QY 509 CCCTGATCTGCACAGGGTACTTCAGGGCTTGTGTCTTTCGAAAAAGCCCGTGGCC 568  
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 262 CTGTGTGCTGCAATGCTCCCTGAGGAGCTCGTTCCTGGGAGATTACCTTGTGCC 321  
QY 569 AACTTGGCGGCGAGGTGTACCAACCTCTGCAATTCAGTGAAGTAGAGAAA 628  
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 322 GGCCCAACAGACCGGGGTGTACACGAACTCTGCAAGTTACCAAGTGTGATCCAGAAA 381  
QY 629 CCGTCCAGNCCA 640  
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 382 CCATCCAGGCCA 393
```

Search completed: September 28, 1999, 11:29:46
Job time: 1867 sec

JOURNAL Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th St SE,
Bothell, WA 98021, USA

FEATURES Location/Qualifiers

source

1.1347
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13"
1.765
/note="androgen regulated"
/product="serine protease"
/codon_start=1
/protein_id="AAD21580.1"
/db_xref="pid:94512030"
/db_xref="gi:4512030"

CDS

translation="MAYGNPWFGLGITLIVAGSLVSGSSQIINEDSPHSQPM
QALVVENELFCSGVLPWPWVLSAHCQNSYITGLHSLADQEPDSQVMSLS
VAHPEYRPLANDLMLIKLDESSEDRIISIASQCPAGNSCLVSGWMLANGR
MPTVLOCVSVSEVCSKLYDPLYPHMPFGAGSGHDQKDSNCNDSGGLICNGYLQ
GIVSFGKAPCGGVPGVYTNCKFTWETKTVQAS"

polyA_site 1327..1332
BASE COUNT 269 a 489 c 334 g 255 t
ORIGIN

Query Match 90.0%; Score 1050; DB 42; Length 1347;
Best Local Similarity 92.4%; Pred. No. 2,4e-230;
Matches 1133; Conservative 0; Mismatches 26; Indels 67; Gaps 1;

QY 2 CGCAGCCCTGCGAGCGGCGACTGTCATGGAACGAAATGTTCTGCTCGGCGTCTG 61
Db 122 CGCAGCCCTGCGAGCGGCGACTGTCATGGAACGAAATGTTCTGCTCGGCGTCTG 181
QY 62 TGCATCCGACGTGGTGTCTGTACGCCGACACTGTTTCCAGAACTCTACACCATCGGCG 121
Db 182 TGCATCCGACGTGGTGTCTGTACGCCGACACTGTTTCCAGAACTCTACACCATCGGCG 241
QY 122 TGGGCGCTGCGACGTGTTGAGGCGGACCAAGAGCGAGGAGCCAGATGTGAGGCGCAGCC 181
Db 242 TGGGCGCTGCGACGTGTTGAGGCGGACCAAGAGCGAGGAGCCAGATGTGAGGCGCAGCC 301
QY 182 TCTCCGTACCGCAGCCAGAGTACAGACACTGTTCTGCTGCTAAGCACTCATCA 241
Db 302 TCTCCGTACCGCAGCCAGAGTACAGACACTGTTCTGCTGCTAAGCACTCATCA 361
QY 242 AGTTGACGAAATCCGTGTCCGAGTGTGACCAATCCGAGCAATCAGATTGCTTCCGAGT 301
Db 362 AGTTGACGAAATCCGTGTCCGAGTGTGACCAATCCGAGCAATCAGATTGCTTCCGAGT 421
QY 302 GCCCTACCGGCGGGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 361
Db 422 GCCCTACCGGCGGGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 481
QY 362 TGGCTACCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 421
Db 482 TGGCTACCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 541
QY 422 TCTATGACCGCGTGTACACCCAGCATGTTTGCCTGCGGCGGAGGCGAAGCAGAGG 481
Db 542 TCTATGACCGCGTGTACACCCAGCATGTTTGCCTGCGGCGGAGGCGAAGCAGAGG 601
QY 482 ACTGCTGCAACGGTGAAGTCTGGGGGGCGCTGANTGCAACGGGTAATTCTCAAGGCGTTG 541
Db 602 ACTGCTGCAACGGTGAAGTCTGGGGGGCGCTGANTGCAACGGGTAATTCTCAAGGCGTTG 661
QY 542 TGTCTTTCGGAAGCCCGCTGTGGCAACTTGGCCAGTGTGCTTACACCACTCT 601
Db 662 TGTCTTTCGGAAGCCCGCTGTGGCAACTTGGCCAGTGTGCTTACACCACTCT 721
QY 602 GCAAAATTCAGTGTGATGAGAGAAACCGTCCAGNCCAGTAACTTGTGGGAGTGGGAA 661
Db 722 GCAAAATTCAGTGTGATGAGAGAAACCGTCCAGGCGCACTTAATCTGTGGGAGTGGGAA 781

QY 662 CCCATGAATGACCCGCCAAATATACCTCGGGAAGATTCAGGAATATCTGTTCCAG 721
Db 782 CCCATGAATGACCCGCCAAATATACCTCGGGAAGATTCAGGAATATCTGTTCCAG 841
QY 722 CCCCTCTCTCTCAGAGCCAGAGTCCAGAGCCCGCCCTCTCTCTCTCAAAACCAAGG 781
Db 842 CCCCTCTCTCTCAGAGCCAGAGTCCAGAGCCCGCCCTCTCTCTCTCAAAACCAAGG 901
QY 782 TACAGATCC----- 791
Db 902 TACAGATCCAGCCCTC 961
QY 791 -----CCAGCCCTC 834
Db 962 CCTCAGACCCAGAGTCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1021
QY 835 CCTCTCTCTCTCAGAGCCAGAGTCCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 894
Db 1022 CCTCTCTCTCTCAGAGCCAGAGTCCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1081
QY 895 CCAGCCCT 954
Db 1082 CCAGCCCT 1141
QY 955 CAGAGGTCCAGCCCGCCCAACCCCTGTTCCCGAGCCAGAGTTCAGGTTCCAGCCCT 1014
Db 1142 CAGAGGTCCAGCCCGCCCAACCCCTGTTCCCGAGCCAGAGTTCAGGTTCCAGCCCT 1201
QY 1015 CTTCTCTCTCTCAGAGCCAGAGTCCAGAGTCCAGAGTTCAGAGTTCAGAGTTCAGAGT 1074
Db 1202 CTTCTCTCTCTCAGAGCCAGAGTCCAGAGTTCAGAGTTCAGAGTTCAGAGTTCAGAGT 1261
QY 1075 GTGGCAGTTCAGCCCACTTACCACTTACCACTTACCACTTACCACTTACCACTTACCACT 1134
Db 1262 GTGGCAGTTCAGCCCACTTACCACTTACCACTTACCACTTACCACTTACCACTTACCACT 1321
QY 1135 CCAGAAATTAAGTTCAGAGAGAGCG 1160
Db 1322 CCAGAAATTAAGTTCAGAGAGAGCG 1347

RESULT 2

AF113141 5900 bp DNA PRI 25-MAR-1999
LOCUS AF113141
DEFINITION Homo sapiens serine protease gene, complete cds.
ACCESSION AF113141
NID 94512031
VERSION AF113141.1 GI:4512031
KEYWORDS
SOURCE
ORGANISM human.
Human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 5900)
AUTHORS Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gellinas,R., Hood,L. and Wang,K.
TITLE Molecular cloning and characterization of prostate, an androgen-regulated serine protease with prostate-restricted expression
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)
MEDLINE 99179024
REFERENCE 2 (bases 1 to 5900)
AUTHORS Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gellinas,R., Hood,L. and Wang,K.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th St SE,
Bothell, WA 98021, USA

FEATURES

source 1..5900
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13"

[illegible]

QY	1024	GACCCACGGGCAATCCACCTAGANNTMCCCTGTACACAGTGGCCCTTGCCAGT	1083
Db	5660	GACCCACGGGTCANATGCCACCTAGACTCTCCCTGTACACAGTGGCCCTTGCCAGT	5719
QY	1084	TGACCCACAGCTTACAGTGGTGTTCATTTTTGTCTTCCTCCCTAGATCAGAAATA	1143
Db	5720	TGACCCACAGCTTACAGTGGTGTTCATTTTTGTCTTCCTCCCTAGATCAGAAATA	5779
QY	1144	AA 1145	
Db	5780	AA 5781	
RESULT	3		
LOCUS	SSU76256		
DEFINITION	Sus scrofa enamel matrix serine proteinase 1 precursor, mRNA,		
ACCESSION	U76256		
NID	G2737920		
VERSION	U76256.1		
KEYWORDS	GI:2737920		
ORGANISM	pig.		
REFERENCE	Sus scrofa		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
TITLE	Eutheria; Cetartiodactyla; Suidae; Sus.		
JOURNAL	1 (bases 1 to 1140)		
REFERENCE	Stimmer, P., Fukae, M., Tanabe, T., Yamakoshi, Y., Uchida, T., Xue, J.,		
AUTHORS	Margolis, H. C., Shimizu, M., Hu, C. C. and Bartlett, J. D.		
TITLE	Purification, Characterization and Cloning of Enamel Matrix Serine		
JOURNAL	Proteinase 1		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 1140)		
TITLE	Stimmer, P., Fukae, M., Tanabe, T., Yamakoshi, Y., Uchida, T., Xue, J.,		
JOURNAL	Margolis, H. C., Shimizu, M., Hu, C. C. and Bartlett, J. D.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (24-OCT-1996) Pediatric Dentistry, University of Texas		
TITLE	Health Science Center at San Antonio, 7703 Floyd Curl Drive, San		
JOURNAL	Antonio, TX 78284-7888, USA		
FEATURES	Location/Qualifiers		
SOURCE	1..1140		
sig_peptide	/organism="Sus scrofa"		
CDS	/db_xref="taxon:9823"		
69..140	/note="the signal peptide is cleaved after Ala24 and the		
69..833	preprotein is secreted into the developing enamel matrix		
/note="EMSP1"			
/codon_start=1			
/product="enamel matrix serine proteinase 1 precursor"			
/protein_id="AAB94638.1"			
/db_xref="PIR:G2737921"			
/db_xref="GI:2737921"			
/translation="MTAAGNPMGWFGLILLGINGSIAFLNGGHIINGEDCNPHSOP			
QAALFIEDFECGVLVHPWLSAHCFOFSYITGLHLDEDPGSGOMDEAS			
IOHPENPSMANDMLIKIKESVISTPTNIVSVCOCPPGSCIVSGWGLASG			
LVQYACNIVISASEKAGKARGPYTHSMFCAGGDDKDCSDGSDGFLICNGSI			
GLVSFQDACCGGPNVGYITNLCFTDIOTTIOAS"			
159..830			
/note="the preprotein is cleaved following His30			
activating the serine proteinase; the active protein has			
an apparent molecular weight of 34 kDa and a derived			
molecular weight of 24 kDa"			
/product="enamel matrix serine proteinase 1"			
/order(177..179,549..551)			
/note="disulfide bond"			
/order(234..236,282..284)			
/note="disulfide bond"			
/join(279..281,414..416,687..689)			
/note="encodes catalytic triad"			
405..407			
/note="encodes potential glycosylation location; yields a			

```

misc_feature      blank cycle during protein sequencing"
order(408..410,789..791)
/Note="disulfide bond"
misc_feature      /note="disulfide bond"
order(510..512,705..707)
/Note="disulfide bond"
misc_feature      /note="disulfide bond"
order(675..677,750..752)
/Note="disulfide bond"
BASE COUNT      245 a 366 c 295 g 234 t
ORIGIN

```

```

Query Match      45.5%; Score 530.6; DB 3; Length 1140;
Best Local Similarity 81.3%; Pred. No. 9.7e-112;
Matches 625; Conservative 0; Mismatches 143; Indels 1; Gaps 1;

QY 2 CGAGAGCCCTGGAGGGGGGAGCTGGTATGGAAGCAATGTTGCTGCGGGGCTCTGG 61
DB 190 CGAGAGCCCTGGAGGGGGGAGCTGGTATGGAAGCAATGTTGCTGCGGGGCTCTGG 249
QY 62 TGCATCCGAGTGGGTGCTGTCAGCCGACACATGTTCCAGAACTCCTACACCATCGGGC 121
DB 250 TGCATCCGAGTGGGTGCTGTCAGCCGACACATGTTCCAGAACTCCTACACCATCGGGC 309
QY 122 TGGGCTTCACAGCTCTTGAAGCCGACCAAGCCAGGAGCCAGATGTTGAGGCCAGCC 181
DB 310 TGGGCTTCACAGCTCTTGAAGCCGACCAAGCCAGGAGCCAGATGTTGAGGCCAGCC 369
QY 182 TCTCGGTAGGGGAGCCAGAGTACAGACACTGTCGCTAGAGCCATGCTCATCA 241
DB 370 TCTCGGTAGGGGAGCCAGAGTACAGACACTGTCGCTAGAGCCATGCTCATCA 429
QY 242 AGTTGAGCAAGTCCGTCGAGTGTGACACATCCGAGACATGCTTCCGAGT 301
DB 430 AGTTGAGCAAGTCCGTCGAGTGTGACACATCCGAGACATGCTTCCGAGT 489
QY 302 GCCCTACCGGGGAACTCTTGGCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 361
DB 490 GCCCTACCGGGGAACTCTTGGCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 549
QY 362 TGCCTACCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 421
DB 550 TGCCTACCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 609
QY 422 TCTATGACCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 481
DB 610 GCTACGGCCCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 669
QY 482 ACTCTGTCAGACGGTGAAGTCTGCGGGGGCCCTGATCTGCAACGGGTACTTGCAGGGCCTTG 541
DB 670 ACTCTGTCAGACGGTGAAGTCTGCGGGGGCCCTGATCTGCAACGGGTACTTGCAGGGCCTTG 729
QY 542 TGTCTTTCGAAAGAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 601
DB 730 TGTCTTTCGAAAGAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 789
QY 602 GCAATTCAGTGAAGTGAAGAGAAACCTGCAAGCCAGTACTTACTTCTGCGGAGTGGGAA 661
DB 790 GCAATTCAGTGAAGTGAAGAGAAACCTGCAAGCCAGTACTTACTTCTGCGGAGTGGGAA 849
QY 662 CCCATGAATTCAGTGAAGTGAAGAGAAACCTGCAAGCCAGTACTTACTTCTGCGGAGTGGGAA 721
DB 850 GCAATTCAGTGAAGTGAAGAGAAACCTGCAAGCCAGTACTTACTTCTGCGGAGTGGGAA 909
QY 722 -GCCCTTCCTCCAGGAGGAGGAGTGCAGGCCGCCAGCCCTTCCTGCC 769
DB 910 TGCCTTCCTCCAGGAGGAGGAGTGCAGGCCGCCAGCCCTTCCTGCC 958

```

```

RESULT 4
AF019979 1237 bp mRNA ROD 01-JAN-1999
LOCUS AF019979
DEFINITION Mus musculus enamel matrix serine proteinase 1 precursor, mRNA,
complete cds.

```

```

ACCESSION AF019979
NID 94090846
VERSION AF019979.1 GI:4090846
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1237)
AUTHORS Simmer, J.
TITLE Enamel Matrix Serine Proteinase 1 (EXSP1)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1237)
AUTHORS Simmer, J.
TITLE Direct Submission
JOURNAL Submitted (18-AUG-1997) Pediatric Dentistry, University of Texas
Health Science Center at San Antonio, 7703 Floyd Curl Drive, San
Antonio, TX 78284-7888, USA
FEATURES
source 1..1237
/organism="Mus musculus"
/strain="Swiss Webster"
/db_xref="taxon:10090"
/cell_type="ameloblast"
sig_peptide 52..126
CDS 52..819
/function="degradation of the enamel matrix during enamel
maturation"
/note="EXSP1"
/codon_start=1
/product="enamel matrix serine proteinase 1 precursor"
/protein_id="AAC98894.1"
/db_xref="pid:94090847"
/db_xref="GI:4090847"
/translation="MAVTRAPWGFUCLILEVTGASASVSRRIIGQDCSPHSP
WQALFSDGFECSGLVHPQWVLSAALQIESIVGLNLKSGQSGRMLEAHL
SIQHNFPDPSFANDLMLIKLNEYSIESNTIRSPVATQCPGPDCLVSGQQLKNG
KLPSLQCVNLSVASEETCRLLYDVEYHLSMFCAGGQDQDSCNGDSGPIVNCNRL
QGLVSGGQKQGPPIPSVYNLCKFTMIQTITN"
misc_feature 127..144
/note="encodes propeptide"
mat_peptide 145..816
/product="matrix serine proteinase 1"
BASE COUNT 289 a 386 c 286 g 276 t
ORIGIN

```

```

Query Match      33.6%; Score 392; DB 12; Length 1237;
Best Local Similarity 65.8%; Pred. No. 4.3e-80;
Matches 628; Conservative 0; Mismatches 311; Indels 16; Gaps 4;

QY 2 CGAGAGCCCTGGAGGGGGGAGCTGGTATGGAAGCAATGTTGCTGCGGGGCTCTGG 61
DB 176 CGAGAGCCCTGGAGGGGGGAGCTGGTATGGAAGCAATGTTGCTGCGGGGCTCTGG 235
QY 62 TGCATCCGAGTGGGTGCTGTCAGCCGACACATGTTCCAGAACTCCTACACCATCGGGC 121
DB 236 TGCATCCGAGTGGGTGCTGTCAGCCGACACATGTTCCAGAACTCCTACACCATCGGGC 295
QY 122 TGGGCTTCACAGCTCTTGAAGCCGACCAAGCCAGGAGCCAGATGTTGAGGCCAGCC 181
DB 296 TGGGCTTCACAGCTCTTGAAGCCGACCAAGCCAGGAGCCAGATGTTGAGGCCAGCC 355
QY 182 TCTCGGTAGGGGAGCCAGAGTACAGACACTGTTGCTGCTAAGCAAGCTCATATATA 241
DB 356 TCTCGGTAGGGGAGCCAGAGTACAGACACTGTTGCTGCTAAGCAAGCTCATATATA 415
QY 242 AGTTGAGCAAGTCCGTCGAGTGTGACACATCCGAGACATGCTTCCGAGT 301
DB 416 AACTGACAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 475
QY 302 GCCCTACCGGGGAACTCTTGGCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 361

```

```

Db 476 GCCCCACTCTGAGATACCTGCTAGTCTCTGTTGGGCTCAACTAAAGATGGAAAC 535
QY 362 TGCTACCGTGTCTGCTAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 421
Db 536 TGCCACCGCTCTCTGCTAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 595
QY 422 TCTATGACCGCTCTGCTAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 481
Db 596 TGTATGACCGCTCTGCTAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 655
QY 482 ACTCTCTGACCGCTCTGCTAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 541
Db 656 ACTCTCTGACCGCTCTGCTAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 715
QY 542 TGTCTCTGACCGCTCTGCTAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 601
Db 716 TGTCTCTGACCGCTCTGCTAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 775
QY 602 GCAATTCAGTGTGAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 661
Db 776 GCAATTCAGTGTGAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 826
QY 662 CCATGAAATGAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 721
Db 827 CTGACTGAAAGGCTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 886
QY 722 CCGCTCTGACCGCTCTGCTAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 781
Db 887 -CCCTCTCTGCTAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 945
QY 782 TACAGATGCTGAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 840
Db 946 CAGA-----CTCAGATTCCTCTCTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 1000
QY 841 CCNTGACAGCCAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 900
Db 1001 AGATCCAGAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 1060
QY 901 CCNTGACAGCCAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 955
Db 1061 CCGCTCTGACCGCTCTGCTAGTGTGAGAGTGTGTTGTGAGGANGTCTGAGTAAGC 1115

RESULT 5
AP135023 4740 bp DNA PRI 20-APR-1999
LOCUS Homo sapiens kallikrein-like protein 1 KLR-L1 gene, partial cds.
DEFINITION AF135023
ACCESSION 94589272
NID AF135023.1 GI:4589272
VERSION AF135023.1 GI:4589272
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 4740)
AUTHORS Yousef, G.M., Luo, L.Y. and Diamond, E.P.
TITLE Identification of novel human kallikrein-like genes on chromosome
19q13.3-q13.4
JOURNAL Biochem. Biophys. Res. Commun. (1999) In press
REFERENCE 2 (bases 1 to 4740)
AUTHORS Yousef, G.M., Luo, L.Y. and Diamond, E.P.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-1999) Pathology and Laboratory Medicine, Mount
Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5,
Canada
FEATURES
source location/Qualifiers
1. 4740
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.3-q13.4"
join("<2263..2425,2847..3097,3181..3317,4588..>4740)

```

```

CDS
/product="kallikrein-like protein 1 KLR-L1"
/locus="2263..2425,2847..3097,3181..3317,4588..4740"
/codon_start=3
/product="kallikrein-like protein 1 KLR-L1"
/protein_id="MAD26424.1"
/db_xref="GI:4589273"
/db_xref="PID:94589273"
/translation="SLVSGSCQIINGEDCSHPQALVMENLFCSGLVHPQW
VISAACFQNSYTGILGLSLDEPDGSDQWVSLVHPVNRPLANDLRLKD
ESVSESDTRSTIASQCPAGNSCTVSGWGLANRMPVLAQVAVSVSEVCSKL
YPLVHPVFCAGGQDQDSCNGBSGPLICNGYLOGIVSFKAPCGVGPVYTN
LCKFEWIKRTVQAS"
BASE COUNT 1106 a 1275 c 1308 g 1051 t
ORIGIN

Query Match 25.2%; Score 293.6; DB 42; Length 4740;
Best Local Similarity 81.4%; Pred. No. 1.2e-57;
Matches 389; Conservative 0; Mismatches 83; Gaps 1;

QY 99 CCAAGTCTCTACACATGCGGCTGCGCTGACAGTCTGAGCGGACAGAGCCAGG 158
Db 2841 CCAAGTCTCTACACATGCGGCTGCGCTGACAGTCTGAGCGGACAGAGCCAGG 2300
QY 159 GACCAAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 218
Db 2901 GACCAAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2960
QY 219 CGCTAACGACCTATGCTATCAAGTGTGAGAGTGTGAGTGTGAGTGTGAGTGTG 278
Db 2961 CGCTAACGACCTATGCTATCAAGTGTGAGAGTGTGAGTGTGAGTGTGAGTGTG 3020
QY 279 GAGCATGACATGCTGTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 338
Db 3021 GAGCATGACATGCTGTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3080
QY 339 GGGTCTGCTGGGCAAC----- 355
Db 3081 GGGTCTGCTGGGCAAC----- 355
QY 355 -----GGCAGATGCTTACCGTCTG 375
Db 3141 TGCCAGATGCTTACCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3200
QY 376 CACTGCTGACATGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 435
Db 3201 CACTGCTGACATGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 3260
QY 436 TACCAACCCAGCATGTTCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 493
Db 3261 TACCAACCCAGCATGTTCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3318

RESULT 6
AF148532 4385 bp DNA PRI 09-JUN-1999
LOCUS Homo sapiens kallikrein 4 (KLK4) gene, complete cds.
DEFINITION AF148532
ACCESSION 95020095
NID AF148532.1 GI:5020095
VERSION AF148532.1 GI:5020095
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 4385)
AUTHORS Stephenson, S.A., Verity, K., Ashworth, L. and Clements, J.A.
TITLE Localization of a new prostate specific antigen-related serine
protease gene, KLK4, is evidence for an expanded human kallikrein
(KLK) gene family cluster on chromosome 19q13.3-13.4
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4385)
AUTHORS Stephenson, S.A., Ashworth, L. and Clements, J.A.

```


QY 1062 ACAGTCCCTTGTGGCAGTTGACCACTTACAGTTGGT 1106
 DB 26977 ACACAGGACACAGGCCCTGTGGAACAAGGATCTAGTTACT 26923

RESULT 10

134189/c 134189 8174 bp DNA PAT 27-JAN-1997
 LOCUS Sequence 5 from patent US 5595900.
 DEFINITION 134189
 ACCESSION 91824980
 NID 134189.1 GI:1824980

VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 TITLE

1 (bases 1 to 8174)
 Lowe, J.B.
 Methods and products for the synthesis of oligosaccharide structures on glycoproteins, glycolipids, or as free molecules, and for the isolation of cloned genetic sequences that determine these structures
 Patent: US 5595900-A 5 21-JAN-1997;
 Location/Qualifiers

FEATURES
 source
 1..8174
 /organism="unknown"

BASE COUNT 1628 a 2229 c 2322 g 1995 t
 ORIGIN

Query Match 15.6%; Score 181.6; DB 5; Length 8174;
 Best Local Similarity 77.3%; Pred. No. 4.2e-32;
 Matches 269; Conservative 0; Mismatches 67; Indels 12; Gaps 4;

QY 700 ATTCAGAAATATGTGTCCAGGCCCTCTCCCTCAGGCCAGAGATCCAGGCCCCAGC 759
 DB 1025 ACTTAGAGATCCAGGCTCCGCGCTCTCTCTCAGACCAAGAGTCCAAAGCCCTGC 966
 QY 760 CCCTCTCCCTCAACCAAGG-GTAGAGATCCAGGCCCTCTCCCTCAGCCAGAG 818
 DB 965 CCTCTCTCTCAGACCCAGAGATCCAGACCCCAAGCCCTCTCTCTCAGCCAGAG 906
 QY 819 TCAGACCCCGCCGCT 870
 DB 905 TCAGA-TCCTTAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 847
 QY 871 CTTTCAGACGAGATCCAGACCCCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 930
 DB 846 CCTCAGACCCAGAGGCCCAAGTCCCGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCT 787
 QY 931 CCCAGCCCT 990
 DB 787 -CCCAACCT 729
 QY 991 CCAGAGTTCAGGTCAGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1038
 DB 728 CCAGAGTTCAGAGCCCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 681

RESULT 11

HUMRSSA19 281 bp DNA PRI 09-MAY-1996
 LOCUS Human 37 bp minisatellite repeats, specific to chromosome 19.
 DEFINITION K03500
 ACCESSION 9337672
 NID K03500.1 GI:337672
 VERSION repeat region.
 KEYWORDS Homo sapiens (clone library: COS 4C) DNA.
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 281)
 Das, H.K., Jackson, C.L., Miller, D.A., Leff, T. and Breslow, J.L.

TITLE The human apolipoprotein C-II gene sequence contains a novel
 JOURNAL chromosome 19-specific minisatellite in its third intron
 MEDLINE J. Biol. Chem. 262 (10), 4787-4793 (1987)
 87165892

FEATURES
 source
 Location/Qualifiers

1..281
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="COS 4C"
 /cell_line="JY"

satellite
 1..37
 /note="sat rpt copy A"

satellite
 38..75
 /note="sat rpt copy B"

satellite
 76..113
 /note="sat rpt copy C"

satellite
 114..151
 /note="sat rpt copy D"

satellite
 152..188
 /note="sat rpt copy E"

satellite
 189..225
 /note="sat rpt copy F"

satellite
 226..262
 /note="sat rpt copy G"

satellite
 263..281
 /note="sat rpt partial copy H"

Query Match 14.7%; Score 171.2; DB 10; Length 281;
 Best Local Similarity 81.2%; Pred. No. 1.3e-29;
 Matches 221; Conservative 0; Mismatches 41; Indels 10; Gaps 2;

QY 781 GTACAGATCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840
 DB 2 GTACAGATCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 61
 QY 841 CCNTGAGCCGAGAGT-----CCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCT 891
 DB 62 CCTCAGACCCAGAGATCAAGACCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 121
 QY 892 CCCCCAGCCCT 951
 DB 122 CCCCCAGCCCT 181
 QY 952 AGTCAGAGTCCCAAGCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1010
 DB 182 ACCAGAGATCCCAAGCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 241
 QY 1011 CCT 1042
 DB 242 CCT 273

RESULT 12

DB7943/c 1665 bp DNA PRI 29-JUL-1997
 LOCUS Human DNA for alpha(1,2)fucosyltransferase, Intron 1.
 DEFINITION DB7943
 ACCESSION 91842173
 NID DB7943.1 GI:1842173
 VERSION Funt1; alpha (1, 2) fucosyltransferase.
 KEYWORDS Homo sapiens peripheral leukocytes DNA.
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Direct Submission
 Submitted (17-SEP-1996) to the DDBJ/EMBL/GenBank databases. Hiroshi
 Kimura, Kurume University, School of Medicine, Department of Legal
 Medicine; 67 Asahimachi, Kurume, Fukuoka 830, Japan

(E-mail: hkimura@kurume.kitarn.or.jp, Tel: 0942-31-7554,
Fax: 0942-31-7700)
2 (sites)
Koda, Y., Soejima, M. and Kimura, H.
Structure and expression of H-type GDP-L-fucose:beta-D-galactoside
2-alpha-L-fucosyltransferase gene (FUT1). Two transcription start
sites and alternative splicing generate several forms of FUT1 mRNA
J. Biol. Chem. 272 (11), 7501-7505 (1997)
97207318
3 (sites)
Larsen, R.D., Ernst, L.K., Nair, R.P. and Lowe, J.B.
Molecular cloning, sequence, and expression of a human
GDP-L-fucose:beta-D-galactoside 2-alpha-L-fucosyltransferase cDNA
that can form the H blood group antigen
Proc. Natl. Acad. Sci. U.S.A. 87 (17), 6674-6678 (1990)
90370848
4 (sites)
Lowe, J.B., Ernst, L.K., Larsen, R.D., Bryant, J.G., Robinson, J.S. and
Kelly, R.J.
Molecular basis for H blood group deficiency in Bombay (Oh) and
para-Bombay individuals
Proc. Natl. Acad. Sci. U.S.A. 91 (13), 5843-5847 (1994)
94286334
Sequence updated (10-Feb-1997) by: Hiroshi Kimura.
Location/Qualifiers
1. 1665
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19q 13.3"
/tissue_type="peripheral leukocytes"
1. 1665
/gene="FUT1"
1. 1665
/gene="FUT1"
/number=1
BASE COUNT 349 a 382 c 513 g 421 t
ORIGIN
Query Match 14.2% Score 165.6; DB: 9; Length 1665;
Best Local Similarity 75.8%; Pred. No. 2.1e-26;
Matches 279; Conservative 0; Mismatches 74; Indels 15; Gaps 6;

HSCPT1175/C
LOCUS 6592 bp DNA PRI 27-APR-1999
DEFINITION Homo sapiens TNNI3 gene.
ACCESSION X80780 X80781 X90782
NID 94071059
KEYWORDS X90780.1 GI:4071059
SOURCE cardiac troponin I; TNNI3 gene.
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 6592)
Baevsar, P.K., Brand, N.J., Yacoub, M.H. and Barton, P.J.R.
Isolation and characterization of the human cardiac troponin I gene
(TNNI3)
Genomics 35 (1), 11-23 (1996)
96299735
2 (bases 1 to 6592)
Barton, P.J.R., Cullen, M.E., Townsend, P.J., Brand, N.J., Mullen, A.J.,
Norman, D.A.M., Baevsar, P.K. and Yacoub, M.H.
Close physical linkage of human troponin genes: organization,
sequence, and expression of the locus encoding cardiac troponin I
and slow skeletal troponin T
Unpublished
3 (bases 1 to 6592)
Baevsar, P.K.
Direct Submission
Submitted (15-AUG-1995) P.K. Baevsar, Cardiothoracic Surgery,
National Heart & Lung Institute, Imperial College of Science, Tech. &
Med., Dovehouse Street, London SW3 6LX, UK
Revised by [3]
4 (bases 1 to 6592)
Barton, P.J.R.
Direct Submission
Submitted (22-DEC-1998) Barton P.J.R., Cardiothoracic Surgery,
National Heart & Lung Institute, Imperial College of Science, Tech. &
Med., Dovehouse Street, London SW3 6LX, UK
On Dec 29, 1998 this sequence version replaced gi:1524065.
This sequence is directly followed by the slow skeletal troponin T
gene (AC A011712, A011713)
Related sequence M64247.
Location/Qualifiers
1. 6592
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone_lib="PI"
/map="P13.3-q13.3"
/map="P13.3-q13.3"
1075. 1081
/note="putative"
join(1107..1260,1532..1544,1706..1789,2195..2236,
2507..2638,4009..4098,4633..4809,6442..6592)
/gene="TNNI3"
1107..6592
/gene="TNNI3"
1107..1249
/gene="TNNI3"
/number=1
prim_transcript 1107..6592
/gene="TNNI3"
1107..1249
5'UTR
CDS
/gene="TNNI3"
join(1250..1260,1532..1544,1706..1789,2195..2236,
2507..2638,4009..4098,4633..4809,6442..6525)
/gene="TNNI3"
/codon_start=1
/product="cardiac troponin I"
/protein_id="CA62301.1"
/db_xref="PID:e196058"
/db_xref="PID:g1524066"
/db_xref="GI:1524066"
/db_xref="SWISS-PROT:P19429"

RESULT 13


```

repeat_region 7478..7839 /rpt_family="MER50"
repeat_region complement(7869..7920)
repeat_region /rpt_family="(TA)n"
repeat_region 7937..8213 /rpt_family="MER4_Internal"
repeat_region complement(8259..9153)
repeat_region /rpt_family="L1Pa2"
repeat_region 9684..9975
repeat_region /rpt_family="AluDo"
repeat_region 11171..11311
repeat_region /rpt_family="MIR"
repeat_region complement(12081..12261)
repeat_region /rpt_family="MIR"
repeat_region 12362..12397
repeat_region /rpt_family="MLTID"
repeat_region 12599..12691
repeat_region /rpt_family="AluSg"
repeat_region 12692..12869
repeat_region /rpt_family="MLTID"
repeat_region complement(13008..13038)
repeat_region /rpt_family="AluSg"
repeat_region complement(13074..13218)
repeat_region /rpt_family="MER11A"
repeat_region complement(13256..13442)
repeat_region /rpt_family="MER11A"
repeat_region complement(13622..13993)
repeat_region /rpt_family="HERK"
repeat_region complement(14138..14301)
repeat_region /rpt_family="AluSg"
repeat_region complement(14308..14692)
repeat_region /rpt_family="MER57_Internal"
repeat_region 14855..15152
repeat_region /rpt_family="AluSp"
repeat_region complement(15179..16657)
repeat_region /rpt_family="MER5_Internal"
repeat_region complement(16026..16426)
repeat_region /note="D5 similarity to overlapping ESTs:-(16426..16026)
repeat_region W43007 zc24g09.s1 Soares senescent fibroblasts NbHSF Homo
repeat_region sapiens cDNA clone 323296 3' similar to gb:M20882
repeat_region PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN E PRECURSOR
repeat_region (HUMAN); (401..10): 99% identity.-(16494..16238) W49630
repeat_region zc42h12.r1 Soares senescent fibroblasts NbHSF Homo sapiens
repeat_region cDNA clone 325031 5' similar to gb:M3575
repeat_region PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN C PRECURSOR
repeat_region (HUMAN); (268..524): 99% identity."
repeat_region complement(16667..16961)
repeat_region /rpt_family="AluY"
repeat_region complement(16963..17056)
repeat_region /rpt_family="MER5_Internal"
repeat_region 17252..17402
repeat_region /rpt_family="MER5B"
repeat_region complement(18346..18385)
repeat_region /note="D5 similarity to W43007 zc24g09.s1 Soares
repeat_region senescent fibroblasts NbHSF Homo sapiens cDNA clone 323296
repeat_region 3' similar to gb:M20882 PREGNANCY-SPECIFIC BETA-1
repeat_region GLYCOPROTEIN E PRECURSOR (HUMAN); (441..402): 100%
repeat_region identity."
repeat_region complement(18346..18429)
repeat_region /note="D5 similarity to W49630 zc42h12.r1 Soares
repeat_region senescent fibroblasts NbHSF Homo sapiens cDNA clone 325031
repeat_region 5' similar to gb:M3575 PREGNANCY-SPECIFIC BETA-1
repeat_region GLYCOPROTEIN C PRECURSOR (HUMAN); (184..267): 100%
repeat_region identity."
repeat_region 18772..19186
repeat_region /rpt_family="L1R16B"
repeat_region complement(19312..19498)
repeat_region /rpt_family="MIR"
repeat_region complement(19577..19678)
repeat_region /rpt_family="MIR"
repeat_region 19679..19978
repeat_region /rpt_family="L1"
repeat_region 19981..20294

```

```

repeat_region /rpt_family="Aluub"
repeat_region 20299..20483
repeat_region /rpt_family="L1"
repeat_region 20467..22397
repeat_region /rpt_family="L1MB5"
repeat_region complement(23264..23758)
repeat_region /note="BLASTX similarity to (331..495): match: 0.89,
repeat_region score: 3.9e-158; database searched: nr; pregnancy-specific
repeat_region beta-1-glycoprotein 11 form r precursor - human"
repeat_region complement(23506..23688)
repeat_region /note="D5 similarity to W49630 zc42h12.r1 Soares
repeat_region senescent fibroblasts NbHSF Homo sapiens cDNA clone 325031
repeat_region 5' similar to gb:M3575 PREGNANCY-SPECIFIC BETA-1
repeat_region GLYCOPROTEIN C PRECURSOR (HUMAN); (1..183): 99% identity."
repeat_region complement(23507..23758)
repeat_region /note="BLASTX similarity to (331..414): match: 0.86,
repeat_region score: 5.7e-104; database searched: nr; pregnancy-specific
repeat_region beta-1-glycoprotein 11 form r precursor - human"
repeat_region complement(23561..23760)
repeat_region /note="BLASTN similarity to T39371 (75..274): match: 1,
repeat_region score: 1.7e-95; database searched: est; ya05b01.r1 Homo
repeat_region sapiens cDNA clone 60533 5' similar to sp:A33514 A33514
repeat_region PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN PRECURSOR -"
repeat_region complement(23685..23760)
repeat_region /note="BLASTN similarity to T39410 (75..150): match: 1,
repeat_region score: 7.0e-47; database searched: est; ya05f02.r3 Homo
repeat_region sapiens cDNA clone 60603 5' contains Alu repetitive
repeat_region element."
repeat_region complement(23704..23760)
repeat_region /note="BLASTN similarity to T49544 (203..259): match:
repeat_region 0.96, score: 1.9e-92; database searched: est; ya76h05.r1
repeat_region Homo sapiens cDNA clone 67641 5' similar to gb:M20881
repeat_region PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN D PRECURSOR"
repeat_region complement(24158..24362)
repeat_region /note="BLASTN similarity to T49544 (1..205): match: 0.98,
repeat_region score: 1.9e-92; database searched: est; ya76h05.r1 Homo
repeat_region sapiens cDNA clone 67641 5' similar to gb:M20881
repeat_region PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN D PRECURSOR"
repeat_region complement(24158..24188)
repeat_region /note="BLASTN similarity to T39410 (47..77): match: 0.96,
repeat_region score: 7.0e-47; database searched: est; ya05f02.r3 Homo
repeat_region sapiens cDNA clone 60603 5' contains Alu repetitive
repeat_region element."
repeat_region complement(24158..24188)
repeat_region /note="BLASTN similarity to T39371 (47..77): match: 0.96,
repeat_region score: 1.7e-95; database searched: est; ya05b01.r1 Homo
repeat_region sapiens cDNA clone 60533 5' similar to sp:A33514 A33514
repeat_region PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN PRECURSOR -"
repeat_region complement(24162..24437)
repeat_region /note="BLASTX similarity to (238..329): match: 0.9,
repeat_region Query Match 14.1%; Score 164; DB 11; Length 205251;
repeat_region Best Local Similarity 80.0%; Pred. No. 3; 6e-28;
repeat_region Matches 240; Conservative 0; Mismatches 48; Indels 12; Gaps 4;
misc_feature misc_feature
misc_feature complement(24158..24188)
misc_feature /note="BLASTN similarity to T39371 (47..77): match: 0.96,
misc_feature score: 1.7e-95; database searched: est; ya05b01.r1 Homo
misc_feature sapiens cDNA clone 60533 5' similar to sp:A33514 A33514
misc_feature PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN PRECURSOR -"
misc_feature complement(24162..24437)
misc_feature /note="BLASTX similarity to (238..329): match: 0.9,
misc_feature Query Match 14.1%; Score 164; DB 11; Length 205251;
misc_feature Best Local Similarity 80.0%; Pred. No. 3; 6e-28;
misc_feature Matches 240; Conservative 0; Mismatches 48; Indels 12; Gaps 4;
QY 717 CCAAGCCCTCTCTCCAGAGCCAGAGTCAGAGCCAGCCCTCTCTCAAC 776
QY 201698 CTAGAGCCCTCTCTCTCAAGAGTCAGAGCCAGCCCTCTCTCAAC 201639
QY 777 AAGG-ATACAGATCCAGAGCCCTCTCTCTCAAGAGTCAGAGCCAGCC 835
QY 201638 CAGAGTCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 201580
QY 836 CTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 887
QY 201579 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 201520
QY 888 CACAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 947
QY 201519 CAG-GCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 201462
QY 948 TCAGAGTCAGAGTCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1007

```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 09:57:00 ; Search time 273 Seconds
(without alignments)
1069.502 Million cell updates/sec

Title: US-09-030-606-175
Perfect score: 1167
Sequence: 1 GCGAGCCCTGCGAGCGGC.....NTAAGAGAGCGCAAAAAA 1167

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1151	98.6	1167	1 V58647	Prostate tumour sp
2	1151	98.6	1167	1 V61252	CDNA sequence of p
3	1111.2	95.2	1248	1 V58644	Prostate tumour sp
4	1111.2	95.2	1248	1 V61249	CDNA sequence of p
5	1036.2	88.8	1265	1 V58645	Prostate tumour sp
6	1036.2	88.8	1265	1 V61250	CDNA sequence of p
7	950.2	81.4	1386	1 V11855	Human sapiens Tub I
8	721.4	61.8	871	1 V37495	Human prostate-spe
9	352.2	30.2	1119	1 V58648	Prostate tumour sp
10	352.2	30.2	1119	1 V61253	CDNA sequence of p
11	331.4	28.4	402	1 X41114	Human secreted pro
12	292.8	25.1	1459	1 V58646	Prostate tumour sp
13	292.8	25.1	1459	1 V61251	CDNA sequence of p
14	228.8	19.6	234	1 V58522	Prostate tumour sp
15	228.8	19.6	234	1 V61168	CDNA sequence of p
16	188.4	16.1	1476	1 X16295	Human Kallikrein e
17	181.6	15.6	8174	1 Q56908	Human alpha(1,2)-f
18	181.6	15.6	8174	1 Q56908	Human alpha(1,2)-f
19	181.6	15.6	8174	1 Q56908	Human alpha(1,2)-f
20	154.6	13.2	1089	1 T39783	Human amyloid prec
21	153	13.1	986	1 Q81203	Human stratum corn
22	143.6	12.3	1438	1 T91126	Human serine prote
23	143.6	12.3	1526	1 V07152	Protease M, a nove
24	140	12.0	732	1 O53487	Human adeno-associ
25	137.4	11.8	4067	1 O53192	Human secreted pro
26	126	10.8	1146	1 V84589	DNA encoding a hum
27	125.4	10.7	833	1 V42925	A. contortrix prot
28	124.6	10.7	693	1 V61861	Human spleen tryps
29	111.8	9.6	741	1 N81653	Human pancreatic t
30	111.8	9.6	744	1 T03999	Human pancreatic t
31	111.8	9.6	744	1 T04000	Human pancreatic t
32	111.8	9.6	744	1 T04001	Human pancreatic t
33	110.6	9.5	1333	1 T48519	Human neuropsin-en
34	110.6	9.5	1333	1 T63251	Mouse neuropsin ge
35	110.2	9.4	790	1 V24548	Trypsinogen-like p
36	107.8	9.2	683	1 O63794	Bovine trypsin gen
37	106.2	9.1	701	1 O63795	Bovine trypsinogen
38	100.8	8.6	897	1 T49878	Porcine trypsinoge
39	99.4	8.5	701	1 X23298	Human TRIT trypsin
40	98	8.4	944	1 V84052	Nucleic acid encod
41	95.8	8.2	957	1 Q20501	Encodes fibrinogen
42	90.6	7.8	1003	1 T13316	Korean Viper Salmo
43	88.2	7.6	925	1 N93196	Monkey recombinant

ALIGNMENTS

RESULT 1					
ID	V58647	standard; CDNA: 1167 BP.			
AC	V58647;				
DT	08-DEC-1998	(first entry)			
DE	Prostate tumour specific gene clone DE13.				
KM	Prostate tumour specific gene; human; prostate cancer; detection;				
OS	therapy; ss.				
FS	Homo sapiens.				
FT	Key	Location/Qualifiers			
FT	CDS	28..645			
FT	FT	/*tag= a			
PD	MO9837418-A2.				
PD	27-AUG-1998.				
PF	25-FEB-1998; U03690.				
PR	09-FEB-1998; US-904809.				
PR	25-FEB-1997; US-806596.				
PR	01-AUG-1997; US-904809.				
PA	(CORI-) CORIXA CORP.				
PI	Dillon DC, Xu J.				
DR	WPI: 98-480805/41.				
DR	P-PSDB: W69388.				
PT	Novel human prostate specific tumour protein and fragments - useful				
PS	for detecting and treating prostate cancers				
PS	Claim 1: Page 115: 14pp; English.				
CC	This sequence represents a human prostate tumour specific gene, and can				
CC	be used in the method of the invention. The method is for detecting				
CC	prostate cancer comprising contacting a biological sample with an agent				
CC	able to bind an immunogenic portion of a prostate protein (such as				
CC	encoded by this sequence). An antibody which binds to an immunogenic				
CC	portion of the prostate protein, and the method can be used to detect,				
CC	monitor progression of, or treat prostate cancers. The antibody may				
CC	also be conjugated to a therapeutic agent for use in therapy of prostate				
CC	cancers.				
SO	Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;				
Query Match	98.6%; Score 1151; DB 1; Length 1167;				
Best Local Similarity	100.0%; Pred. No. 2.7e-278;				
Matches 1167; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
OY	1 GCGAGCCCTGCGAGCGGCATGTCATGGAACGAAATGTTCTGGGGGCTCTG	60			
DB	1 GCGAGCCCTGCGAGCGGCATGTCATGGAACGAAATGTTCTCTGGGGGCTCTG	60			
OY	61 GTGCATCCGAGTGGTGTCTGTCAGCCGACACTGTTCCAGAACTCTACACCATCGGG	120			
DB	61 GTGCATCCGAGTGGTGTCTGTCAGCCGACACTGTTCCAGAACTCTACACCATCGGG	120			
OY	121 CTGGGCTGCGACACTGTTGAGGCCGACCAAGAGCCAGGAGCCAGATGTGGAGCCAGC	180			
DB	121 CTGGGCTGCGACACTGTTGAGGCCGACCAAGAGCCAGGAGCCAGATGTGGAGCCAGC	180			
OY	121 CTGGGCTGCGACACTGTTGAGGCCGACCAAGAGCCAGGAGCCAGATGTGGAGCCAGC	180			
DB	121 CTGGGCTGCGACACTGTTGAGGCCGACCAAGAGCCAGGAGCCAGATGTGGAGCCAGC	180			
OY	181 CTCGCGTAGGCGACCAAGATCAACAGACTCTGCTGCTAAGCACTCATGCTCATC	240			
DB	181 CTCGCGTAGGCGACCAAGATCAACAGACTCTGCTGCTAAGCACTCATGCTCATC	240			
OY	181 CTCGCGTAGGCGACCAAGATCAACAGACTCTGCTGCTAAGCACTCATGCTCATC	240			
DB	181 CTCGCGTAGGCGACCAAGATCAACAGACTCTGCTGCTAAGCACTCATGCTCATC	240			
OY	241 AAGTGGACGATCCGTCGATCGATCGACATCCGAGCATCAGCATTTGCTTCGAG	300			
DB	241 AAGTGGACGATCCGTCGATCGATCGATCGACATCCGAGCATCAGCATTTGCTTCGAG	300			
OY	301 TGCCCTACCGCGGGGAACTCTGGCTGCTGCTGGGGTCTGTCGCGAAGCGGAGA	360			
DB	301 TGCCCTACCGCGGGGAACTCTGGCTGCTGCTGGGGTCTGTCGCGAAGCGGAGA	360			
OY	361 ATGCTACCGGTGTCGACATGCGGTGAACGTGTGGTGTGTGAGGANGTCTGAGTAAG	420			
DB	361 ATGCTACCGGTGTCGACATGCGGTGAACGTGTGGTGTGTGAGGANGTCTGAGTAAG	420			

Db	361	ATGCGTACCGGTGCGTCACTCGTGGTAACGTGTGCGTGTCTCTGAGSANGTCTGCAGTAA	420
QY	421	CTCTATGACCGCGTGTACCAACCCAGCATGTTCTGCGCGCGGAGGGCAGACCAAG	480
Db	421	CTCTATGACCGCGTGTACCAACCCAGCATGTTCTGCGCGCGGAGGGCAGACCAAG	480
QY	481	GACTCCGCAACGGTGTACTGTGGGGGGCCCGTATGTGCAAGGGGTACTTGCAGGGCCTT	540
Db	481	GACTCCGCAACGGTGTACTGTGGGGGGCCCGTATGTGCAAGGGGTACTTGCAGGGCCTT	540
QY	541	GTGCTTTCCGAAAAAGCCCGGTGTGGCCAACTTGGCGGTCCAGTGTCTACACCACTC	600
Db	541	GTGCTTTCCGAAAAAGCCCGGTGTGGCCAACTTGGCGGTCCAGTGTCTACACCACTC	600
QY	601	TGCAAAATTCAGTGTGATAGAAAAACCGTCACAGNCAGTTAACTGTGGGACTGGGA	660
Db	601	TGCAAAATTCAGTGTGATAGAAAAACCGTCACAGNCAGTTAACTGTGGGACTGGGA	660
QY	661	ACCCATGAATTTGACCCCCCAATTCATCCTGTGGGAAANGAATTCAGGAATATCTGTCCA	720
Db	661	ACCCATGAATTTGACCCCCCAATTCATCCTGTGGGAAANGAATTCAGGAATATCTGTCCA	720
QY	721	GCCCTCTCTCCCTCAGGCGCCAGAGTCCAGGCCCCCAGCCCTCTCTCTCTCAAACCAAG	780
Db	721	GCCCTCTCTCCCTCAGGCGCCAGAGTCCAGGCCCCCAGCCCTCTCTCTCAAACCAAG	780
QY	781	GTACAGATCCCCAGCCCTCTCTCTCTCAGACCCAGGAGTCCAGACCCCCAGCCCTCT	840
Db	781	GTACAGATCCCCAGCCCTCTCTCTCTCAGACCCAGGAGTCCAGACCCCCAGCCCTCT	840
QY	841	CCNTCAGACCCAGGAGTCCAGCCCTCTCTCTCAGACGAGGAGTCCAGACCCCCAGC	900
Db	841	CCNTCAGACCCAGGAGTCCAGCCCTCTCTCTCAGACGAGGAGTCCAGACCCCCAGC	900
QY	901	CCNTCNTCCGTCAAGACCCAGGGGTGCAAGCCCCCAACCCCTCTCAGAGTATAGAGG	960
Db	901	CCNTCNTCCGTCAAGACCCAGGGGTGCAAGCCCCCAACCCCTCTCAGAGTATAGAGG	960
QY	961	TCCAAAGCCCCCAACCCCTCTCTCTCTCCAGACCCAGAGTTCAGGTCCAGCCCCCTCTCC	1020
Db	961	TCCAAAGCCCCCAACCCCTCTCTCTCTCCAGACCCAGAGTTCAGGTCCAGCCCCCTCTCC	1020
QY	1021	TCAGACCCAGGCGTCCATGCAATGCACACTAGANTTCCCTGTACACAGTCCCCCTTGTGCA	1080
Db	1021	TCAGACCCAGGCGTCCATGCAATGCACACTAGANTTCCCTGTACACAGTCCCCCTTGTGCA	1080
QY	1081	NGTTGACCCCAACTTACCAAGTGTGGTTTTCATTTTGTCCCTTCCCTAGATCCAGAA	1140
Db	1081	NGTTGACCCCAACTTACCAAGTGTGGTTTTCATTTTGTCCCTTCCCTAGATCCAGAA	1140
QY	1141	ATTAAGTNTAAGAGAACGCGAAAAAA 1167	
Db	1141	ATTAAGTNTAAGAGAACGCGAAAAAA 1167	

RESULT	2
ID	V61252 standard; cDNA; 1167 BP.
AC	V61252;
DT	06-JAN-1999 (first entry)
DE	cDNA sequence of prostate tumour clone P703 splice variant DEL3.
KW	Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS	Homo sapiens.
PN	W09837093-A2.
PD	27-AUG-1998.
PF	25-FEB-1998; U03492.
PR	09-FEB-1998; US-020956.
PR	25-FEB-1997; US-806099.
PR	01-AUG-1997; US-904804.
PA	(CORI-) CORIAXA CORP.
PI	Dillon DC, Xu J.
DR	MP1; 98-A09886/51.

DR P-PSDB:W71782.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PS used in a vaccine for the treatment of prostate cancer
CS Claim 3, Page 107; 130pp; English
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;

Query Match	98.6%;	Score 1151;	DB 1;	length 1167;
Best Local Similarity	100.0%;	Pred. No. 2.7e-278;		
Matches 1167; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	1	GGGAGCCCTGGCAGGCGGCACTGGTCAATGAAAAAGATTGTTTCGTGCGGGGCGTCTG	60
Db	1	GGGAGCCCTGGCAGGCGGCACTGGTCAATGAAAAAGATTGTTTCGTGCGGGGCGTCTG	60
QY	61	GTGCATCCGCAATGGGTGCTGTACGCGGCACACTGTTTCCAAACCTCTACACATCGGG	120
Db	61	GTGCATCCGCAATGGGTGCTGTACGCGGCACACTGTTTCCAAACCTCTACACATCGGG	120
QY	121	CTGGGCTTCACACAGTCTTATGAGCCGACACAGAGCCAGAGGACCATGTTGGAGGCCACG	180
Db	121	CTGGGCTTCACACAGTCTTATGAGCCGACACAGAGGACCATGTTGGAGGCCACAG	180
QY	181	CTCTCGTAGCGGACCCACAGATACAAACAGACTTGTGCTATACAGACCTCATGTCATC	240
Db	181	CTCTCGTAGCGGACCCACAGATACAAACAGACTTGTGCTATACAGACCTCATGTCATC	240
QY	241	AAGTTGGAGCAATCCGTGTGCGAGTGTGACACACATCCGAGGACATCAGATTGCTTGGCAG	300
Db	241	AAGTTGGAGCAATCCGTGTGCGAGTGTGACACACATCCGAGGACATCAGATTGCTTGGCAG	300
QY	301	TGCCTTACCGGGGGGAACTTGTCTCGTNTCTGGTGGGGTCTGTGGCGAAACGGCAAA	360
Db	301	TGCCTTACCGGGGGGAACTTGTCTCGTNTCTGGTGGGGTCTGTGGCGAAACGGCAAA	360
QY	361	ATGCGTACCGTCTCTGACACTGCGGTGAACGTCGCGTGTCTGAGGANTCTGCAGTAAG	420
Db	361	ATGCGTACCGTCTCTGACACTGCGGTGAACGTCGCGTGTCTGAGGANTCTGCAGTAAG	420
QY	421	CTCTATGACCGCGCTGTACCAACCCACAGCATGTTCTGCGCGGGGAGGCGAAGCCGAAG	480
Db	421	CTCTATGACCGCGCTGTACCAACCCACAGCATGTTCTGCGCGGGGAGGCGAAGCCGAAG	480
QY	481	GACTCTGCAAGGATGACTCTGGGGGGCCCGATGTGAAGGGATCTTGCAGGGCCCTT	540
Db	481	GACTCTGCAAGGATGACTCTGGGGGGCCCGATGTGAAGGGATCTTGCAGGGCCCTT	540
QY	541	GTGTCTTTTGGAAAAAGCCCGTGTGGCCAACTTGGGCTCCAGGTTGTACACCAACTTC	600
Db	541	GTGTCTTTTGGAAAAAGCCCGTGTGGCCAACTTGGGCTCCAGGTTGTACACCAACTTC	600
QY	601	TGCAAAATTACTGATGATGATAGAGAAAACCGTTCAGNCCAGTTAACTCTGGGGACTGGGA	660
Db	601	TGCAAAATTACTGATGATGATAGAGAAAACCGTTCAGNCCAGTTAACTCTGGGGACTGGGA	660
QY	661	ACCCATGAAATTAAGCCCCCAATACTCTGTGGGAAGAAATTCAGAAATATCTGTTCCCA	720
Db	661	ACCCATGAAATTAAGCCCCCAATACTCTGTGGGAAGAAATTCAGAAATATCTGTTCCCA	720
QY	721	GGCCCTCTCTCCCTCAGGGCCAGAGTCAAGGGCCCGCAGCCCTCTCTCCCTCAAACTCAAG	780
Db	721	GGCCCTCTCTCCCTCAGGGCCAGAGTCAAGGGCCCGCAGCCCTCTCTCCCTCAAACTCAAG	780
QY	781	GTACAGATCCCAAGCCCTCTCTCTCTCAGACCCAGAGTCCAGACCCCGCAGCCCTCTCT	840
Db	781	GTACAGATCCCAAGCCCTCTCTCTCTCAGACCCAGAGTCCAGACCCCGCAGCCCTCTCT	840

Sequence	1248 BP	288 A	424 C	303 G	228 T
QY	841	CCCTGAGACCCAGAGAGTCCACCCCTCTCTCTCTGAGAGAGAGAGTCCACACCCCCAGC	900		
Db	841	CCNTGAGACCCAGAGAGTCCACCCCTCTCTCTGAGAGAGAGAGTCCACACCCCCAGC	900		
QY	901	CCNTCTCCCTGAG	960		
Db	901	CCNTCTCCCTGAG	960		
QY	961	TCCAAGCCCCCAACCCCT	1020		
Db	961	TCCAAGCCCCCAACCCCT	1020		
QY	1021	TCAGACCCAG	1080		
Db	1021	TCAGACCCAG	1080		
QY	1081	NGTTGACCCCAACCTTACCACTTGGTTTTCATTTTTTTTGTCCCTTCCCTCAGATCAGAA	1140		
Db	1081	NGTTGACCCCAACCTTACCACTTGGTTTTCATTTTTTTTGTCCCTTCCCTCAGATCAGAA	1140		
QY	1141	ATTAAGTTTAAG	1167		
Db	1141	ATTAAGTTTAAG	1167		
RESULT	3				
V58644					
AC	V58644 standard; cDNA: 1248 BP.				
DT	V58644:				
DE	08-DEC-1998 (first entry)				
KW	Prostate tumour specific gene clone DEL.				
OS	Prostate tumour specific gene; human; prostate cancer; detection;				
FT	Key	Location/Qualifiers			
FT	CDS	217..696			
FT		/*tag= a			
PN	W09837418-A2.				
PD	27-AUG-1998.				
PF	25-FEB-1998; U03690.				
PR	09-FEB-1998; US-904809.				
PR	25-FEB-1997; US-806596.				
PR	01-AUG-1997; US-904809.				
PA	(CORI-) CORIXA CORP.				
PI	Dillon DC, Xu J.				
DR	WPI; 98-480805/41.				
DR	P-PSDB; W69387.				
PT	Novel human prostate specific tumour protein and fragments - useful				
PT	for detecting and treating prostate cancers				
PS	Claim 1; Page 112; 14pp; English.				
CC	This sequence represents a human prostate tumour specific gene, and can				
CC	be used in the method of the invention. The method is for detecting				
CC	prostate cancer comprises contacting a biological sample with an agent				
CC	able to bind an immunogenic portion of a prostate protein (such as				
CC	encoded by this sequence). An antibody which binds to an immunogenic				
CC	portion of the prostate protein, and the method can be used to detect,				
CC	monitor progression of, or treat prostate cancers. The antibody may				
CC	also be conjugated to a therapeutic agent for use in therapy of prostate				
CC	cancers.				

	Query Match	95.28;	Score 1111.2;	DB 1;	Length 1248;
	Best Local Similarity	96.7%;	Pred. No. 2,3e-268;		
	Matches 1139;	Conservative	0;	Mismatches 27;	Indels 12;
				Gaps	1
QY	2 CGCAGCCCTGGCAGGCGGCACCTGTCATGGAAGCAATGTTCTGCTCGGGCGCTCTGG	61			
DB	41 CGCAGCCCTGGCAGGCGGCACCTGTCATGGAAGCAATGTTCTGCTCGGGCGCTCTGG	100			
QY	62 TGCATCCGAGTGGGTGCTGTATGACCGCGACACTGTTCCAGAA-----CTCCT	109			
DB	101 TGCATCCGAGTGGGTGCTGTATGACCGCGACACTGTTCCAGAAAGTGAAGCAGAGCTCT	160			

QY	110	ACACCATGSGGCTG6GGCTTCACAGTCTTGAGGCGGACCAACAGCCAGGAGCCAGATGG	165
Db	161	ACACCATGSGGCTG6GGCTTCACAGTCTTGAGGCGGACCAACAGCCAGGAGCCAGATGG	220
QY	170	TGAGAGCAGGCTCTCCGTACAGGGCACGAGAGTACAAAGACTCTTGCTCCGTAAAGACC	229
Db	221	TGAGAGCAGGCTCTCCGTACAGGGCACGAGAGTACAAAGACTCTTGCTCCGTAAAGACC	280
QY	230	TCATGCTCATCAAGTTGGAGGAATCCGTTGCGAGTCTGACACATCCGGAGCATCAGA	289
Db	281	TCATGCTCATCAAGTTGGAGGAATCCGTTGCGAGTCTGACACATCCGGAGCATCAGA	340
QY	290	TTTGCTTGCGAGTGGCTTACCCGGGGGAACTCTTGCTCGTNTCTGGCTG6G6TCTG	349
Db	341	TTTGCTTGCGAGTGGCTTACCCGGGGGAACTCTTGCTCGTNTCTGGCTG6G6TCTG	400
QY	350	CGAAGCGAGAAATGCGACGCTGCTACGTGGTGAAGCTGCGGTGGTGTAGGANG	409
Db	401	CGAAGCGAGAAATGCGACGCTGCTACGTGGTGAAGCTGCGGTGGTGTAGGANG	460
QY	410	TCCTCAGTAAGCTCTATAGACCCGCTTACACCCAGCATGTTCTGCGCCCGGAGAGGC	469
Db	461	TCCTCAGTAAGCTCTATAGACCCGCTTACACCCAGCATGTTCTGCGCCCGGAGAGGC	520
QY	470	AAGACCAAGAAAGACTCTCTGCAAGGTAAGTCTG6GGGGGCCCCGTGATCTGCAACGGGTACT	529
Db	521	AAGACCAAGAAAGACTCTCTGCAAGGTAAGTCTG6GGGGGCCCCGTGATCTGCAACGGGTACT	580
QY	530	TGCAAGGCTTGTGTCTTCTTG6GAAAAAGCCCGTGTG6CCAATTGGCGTGGCAGGTGTCT	589
Db	581	TGCAAGGCTTGTGTCTTCTTG6GAAAAAGCCCGTGTG6CCAATTGGCGTGGCAGGTGTCT	640
QY	590	ACACCAACCTCTGCAATTTACTAGAGGGAATAGAGAAAAAGCTGCAAGNCAAGTTAACTCT	649
Db	641	ACACCAACCTCTGCAATTTACTAGAGGGAATAGAGAAAAAGCTGCAAGNCAAGTTAACTCT	700
QY	650	G6GGACTGGGAAGCCATGAATTTGACCCCAATATACATCTCTCGGANGAATTCAGGAT	709
Db	701	G6GGACTGGGAAGCCATGAATTTGACCCCAATATACATCTCTCGGANGAATTCAGGAT	760
QY	710	ATCTGTTCCAGCCCTCTCTCTCAAGGCCAGAGSTCCAGSCCCCAAGCCCTCTCTCC	769
Db	761	ATCTGTTCCAGCCCTCTCTCTCTCAAGGCCAGAGSTCCAGSCCCCAAGCCCTCTCTCC	820
QY	770	TCAAACCAAGGTTCAATATCCCAAGCCCTCTCTCTCTCAAGCCAGAGSTCCAGACCCCC	829
Db	821	TCAAACCAAGGTTCAATATCCCAAGCCCTCTCTCTCTCTCAAGCCAGAGSTCCAGACCCCC	880
QY	830	CAGCCCTCTCTCTCTCAAGCCAGAGSTCCAGCCCTCTCTCTCTCTCAAGCCAGAGSTCC	889
Db	881	CAGCCCTCTCTCTCTCTCAAGCCAGAGSTCCAGCCCTCTCTCTCTCTCAAGCCAGAGSTCC	940
QY	890	GACCCCCAGCCCTCTCTCTCTCAAGCCAGAGG6GTGCAAGCCCAACCCCTCTCTCTCT	949
Db	941	GACCCCCAGCCCTCTCTCTCTCTCTCAAGCCAGAGG6GTGCAAGCCCAACCCCTCTCTCTCT	1000
QY	950	AGAGTCAGAGSTCCAGACCCCAACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1009
Db	1001	AGAGTCAGAGSTCCAGACCCCAACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1060
QY	1010	CCCCCT	1069
Db	1061	CCCCCT	1120
QY	1070	CCCTTGTGGCAGGTGACCCCAACCTTACAGAGTTG6TTTTTCAATTTTTTGTCCCTTCCC	1129
Db	1121	CCCTTGTGGCAGGTGACCCCAACCTTACAGAGTTG6TTTTTCAATTTTTTGTCCCTTCCC	1180
QY	1130	TAGATCCAGAAATTAAGTNTTAAGAGAGGCAAAAAA	1167
Db	1181	TAGATCCAGAAATTAAGTNTTAAGAGAGGCAAAAAA	1218

```

RESULT 4
ID V61249 standard; cDNA; 1248 BP.
AC V61249;
DE 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE1.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillion DC, Xu J;
DR MPI; 98-609886/51.
DR P-PDB; W71871.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 104; 130bp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 95.2%; Score 1111.2; DB 1; Length 1248;
Best Local Similarity 96.7%; Pred. No. 2.3e-268;
Matches 1139; Conservative 0; Mismatches 27; Indels 12; Gaps 1;

QY 2 CGAGCCCTTGGCAGGCGGCACTGTATGGAAGAAACGATTGCTGCTGGGCGCTCTGG 61
DB 41 CGAGCCCTTGGCAGGCGGCACTGTATGGAAGAAACGATTGCTGCTGGGCGCTCTGG 100
QY 62 TGCATCCGAGTGGGCTCTGTCAGCCGACACTGTTTCCAGAA-----CTCCT 109
DB 101 TGCATCCGAGTGGGCTCTGTCAGCCGACACTGTTTCCAGAA-----CTCCT 160
QY 110 ACACCATCGGCTGGGCTCTGTCAGAGTGTAGAGCCGACAGAGCCAGGAGCCAGATGG 169
DB 161 ACACCATCGGCTGGGCTCTGTCAGAGTGTAGAGCCGACAGAGCCAGGAGCCAGATGG 220
QY 170 TGAAGCCAGCTCTCTCCGTAGGCGACCCAGAGTAAACAGACTTCTGCTAGAGACC 229
DB 221 TGAAGCCAGCTCTCTCCGTAGGCGACCCAGAGTAAACAGACTTCTGCTAGAGACC 280
QY 230 TCATGCTCATCAAGTTGGAGGAATCCGTCCGAGTGTGACACATCCGAGGATCGCA 289
DB 281 TCATGCTCATCAAGTTGGAGGAATCCGTCCGAGTGTGACACATCCGAGGATCGCA 340
QY 290 TTGCTTCGAGTGGCCCTTACCGCGGGGAACTTTCCTCTGTTTCGCTGGGCTCTGCTGG 349
DB 341 TTGCTTCGAGTGGCCCTTACCGCGGGGAACTTTCCTCTGTTTCGCTGGGCTCTGCTGG 400
QY 350 CGAAGCGCAAAATGCTTACCGTGTGTCAGTGGTGAAGCGTGTGGTGTCTGAGGANG 409
DB 401 CGAAGCGCAAAATGCTTACCGTGTGTCAGTGGTGAAGCGTGTGGTGTCTGAGGANG 460
QY 410 TCAGAGTAAGCTCTATGACCCGCTGACACCCACATGTTCTGGCGGGGAGAGGC 469
DB 461 TCAGAGTAAGCTCTATGACCCGCTGACACCCACATGTTCTGGCGGGGAGAGGC 520
QY 470 AAGACCGAAGAGACTCTGCAACGCTGACTCTGGGGGGCCCTGATGTGCAACGGTACT 529
DB 521 AAGACCGAAGAGACTCTGCAACGCTGACTCTGGGGGGCCCTGATGTGCAACGGTACT 580
QY 530 TGCAGGCGCTTGTGTCTTTTGGAAAGCCCGGTGTGCCCAACTTGGCGTCCAGGTCT 589

```

```

DB 581 TGCAGGCGCTTGTGTCTTTGGAAGAGCCCGTGTGCCAGTTGGCGTCCAGGTGTCT 640
QY 590 ACACCAACCTTGCATAATCTACTAGTGTATGAGAAAGCGTCCAGNCAGTTAACTCT 649
DB 641 ACACCAACCTTGCATAATCTACTAGTGTATGAGAAAGCGTCCAGNCAGTTAACTCT 700
QY 650 GGGGACTGGGAACCCATGAATTTGACCCCAATATCATCTCTGGGAANGAATTCAGGAAT 709
DB 701 GGGGACTGGGAACCCATGAATTTGACCCCAATATCATCTCTGGGAANGAATTCAGGAAT 760
QY 710 ATCTGTCCAGAGCCCTCTCTCTCTAGGCGCCAGAGTCTCAGGCGCCAGCCCTCTCTCC 769
DB 761 ATCTGTCCAGAGCCCTCTCTCTCTAGGCGCCAGAGTCTCAGGCGCCAGCCCTCTCTCC 820
QY 770 TCAAAACCAAGGATAGATCCGAGCCCTCTCTCTCTAGGCGCCAGAGTCTCAGAGCCCC 829
DB 821 TCAAAACCAAGGATAGATCCGAGCCCTCTCTCTCTAGGCGCCAGAGTCTCAGAGCCCC 880
QY 830 CAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 889
DB 881 CAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 940
QY 890 GACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 949
DB 941 GACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1000
QY 950 AGAGTGAAGGTGCAAGCCCGCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1009
DB 1001 AGAGTGAAGGTGCAAGCCCGCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1060
QY 1010 CCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1069
DB 1061 CCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1120
QY 1070 CCTTGTGCGAGTTTACCCAACTTACCAAGTTGTTTTCATTTTTCCTCTCTCTCTCT 1129
DB 1121 CCTTGTGCGAGTTTACCCAACTTACCAAGTTGTTTTCATTTTTCCTCTCTCTCTCT 1180
QY 1130 TAGATCCAGAAATTAAGTNTAAGAGAGGCAAAAAA 1167
DB 1181 TAGATCCAGAAATTAAGTNTAAGAGAGGCAAAAAA 1218

RESULT 5
ID V58645 standard; cDNA; 1265 BP.
AC V58645;
DE 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE2.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
PN WO9837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillion DC, Xu J;
DR MPI; 98-480805/41.
PT Novel human prostate specific immunogen and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 113-114; 141bp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.

```

Seq	Sequence	1265 BP	256 A	432 C	321 G	245 T
QY	Query Match	88.8%	Score 1036.2	DB 1	Length 1265	
Best Local Similarity	91.0%	Pred. No. 1.2e-249				
Matches 1137	Conservative	0	Mismatches 29	Indels 83	Gaps	1
QY	2	CGCAGCCCTGGAGGCGGCACCTGTGTCATGTGGAAAAAGAAATTTGTTCTGTGGGCGTCTTG	61			
Db	14	CGCAGCCCTGGAGGCGGCACCTGTGTCATGTGGAAAAAGAAATTTGTTCTGTGGGCGTCTTG	73			
QY	62	TGCATCCGAGTGGGGTGGGTGAGGCGGACACTGTTTCCAGAACTCTACCATATCGGCG	121			
Db	74	TGCATCCGAGTGGGGTGGGTGAGGCGGACACTGTTTCCAGAACTCTACCATATCGGCG	133			
QY	122	TGGGCTGACACAGTCTTGAGGCGGACCAAGAGCCAGGAGGACAGATGTGTGAGGCGGAC	181			
Db	134	TGGGCTGACACAGTCTTGAGGCGGACCAAGAGCCAGGAGGACAGATGTGTGAGGCGGAC	193			
QY	182	TCTCCGTACGGCACCACAGTACAAACAGACTTTGCTGCTTACAGACTCATGCTCATCA	241			
Db	194	TCTCCGTACGGCACCACAGTACAAACAGACTTTGCTGCTTACAGACTCATGCTCATCA	253			
QY	242	AGTTGAGAGAAATCCGTGTCGAGTGCATCCAGCATCCGAGCATCAAGTATGTTGCACT	301			
Db	254	AGTTGAGAGAAATCCGTGTCGAGTGCATCCAGCATCCGAGCATCAAGTATGTTGCACT	313			
QY	302	GCCCTACCCGCGGGGAACTCTTGCTGCTGTTCTGGGTGGGTTGCTGGCGGAC	355			
Db	314	GCCCTACCCGCGGGGAACTCTTGCTGCTGTTCTGGGTGGGTTGCTGGCGGAC	373			
QY	355	-----	355			
Db	374	TCACGGGTGTGTGTGTCGCCCTCTTCAAGAGGTCCCTGACCAGTGCAGCGGGGCTGACCC	433			
QY	355	-----GGAGAAATCCCTACCGTGCATCCGTCAGTCCGTAAGTTCGCTGT	398			
Db	434	AGAGCTTGCGCTCCAGGACAGATGCTTACCTGCTGCGAGTCTGTGTAAGTTCGCTGT	493			
QY	399	GTCGTAGAGANGTCTGTCAGTAAAGCTTATGACCCGCTGTACCAACCCAGCATTTGCGC	458			
Db	494	GTCGTAGAGANGTCTGTCAGTAAAGCTTATGACCCGCTGTACCAACCCAGCATTTGCGC	553			
QY	459	CGCGGAGGGGACAGACCAAGAAAGACTCCTGCAAGGTGACTCTGGGGGGCCCTGATCTG	518			
Db	554	CGCGGAGGGGACAGACCAAGAAAGACTCCTGCAAGGTGACTCTGGGGGGCCCTGATCTG	613			
QY	519	CAAGGGGTACTTGCAGGGGCTTGTCTTTTGGAAAAAGCCCGTGTGGCAACTTGGCGT	578			
Db	614	CAAGGGGTACTTGCAGGGGCTTGTCTTTTGGAAAAAGCCCGTGTGGCAAGTGGCGT	673			
QY	579	GCCAGGTGTCTACACCAACCTCTGCAAAATTCAGTGTGATAGAGAAAAACGCTCCAGNC	638			
Db	674	GCCAGGTGTCTACACCAACCTCTGCAAAATTCAGTGTGATAGAGAAAAACGCTCCAGNC	733			
QY	639	CAGTTAACTGTGGGACTTGGGAACCCATGAATTGACCCCAATATCATCTTGGGAANG	698			
Db	734	CAGTTAACTGTGGGACTTGGGAACCCATGAATTGACCCCAATATCATCTTGGGAANG	793			
QY	699	AATTCAGGAATATCTGTTCCAGAGCCCTCTCTCCATGAGCCCAAGAGTCCAGAGCCCCAG	758			
Db	794	AATTCAGGAATATCTGTTCCAGAGCCCTCTCTCCATGAGCCCAAGAGTCCAGAGCCCCAG	853			
QY	759	CCCCCTCTCCCTCAAAACCAAGGATACAGATCCCAAGCCCTCTCTCTCAGACCCAGAG	818			
Db	854	CCCCCTCTCTCTCAAAACCAAGGATACAGATCCCAAGCCCTCTCTCTCAGACCCAGAG	913			
QY	819	TCACAGACCCCCAGCCCT	878			
Db	914	TCACAGACCCCCAGCCCT	973			
QY	879	GCAGAGATTCAGACCCCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	938			

[illegible]

[illegible]

Query Match	Best Local Similarity	Matches	725; Conservative	0; Mismatches	13; Indels	0; Gaps	0;
Query 2	CGCAGCCTTGGCAGGCGGCACCTGTCATGAGAAACGAATTTGTTGCTCGGGCGTCTCGG	61	100%	0	0	0	0
Db 134	CCGACCCCTGGCAGGCGGCACCTGTCATGAGAAACGAATTTGTTGCTCGGGCGTCTCGG	193	100%	0	0	0	0
Query 62	TGCATCCGCGAGTGGGTGCTGCAGCCGCGACACTGTTTCCAGAACTCTACACATCGGGC	121	100%	0	0	0	0
Db 134	CCGACCCCTGGCAGGCGGCACCTGTCATGAGAAACGAATTTGTTGCTCGGGCGTCTCGG	193	100%	0	0	0	0
Query 134	TGCATCCGCGAGTGGGTGCTGCAGCCGCGACACTGTTTCCAGAACTCTACACATCGGGC	253	100%	0	0	0	0
Db 122	TGGGCTCTGACAGTCTTGGAGCCGACCAAGAGCCAGGAGCCAGATGGTGGAGGCCACCC	181	100%	0	0	0	0
Db 254	TGGGCTCTGACAGTCTTGGAGCCGACCAAGAGCCAGGAGCCAGATGGTGGAGGCCACCC	313	100%	0	0	0	0
Query 132	TCTCCGTACGGCAGCCAGATCAACAGACTTTGCTGCTTACAGCACTCATGCTCATCA	241	100%	0	0	0	0
Db 314	TCTCCGTACGGCAGCCAGATCAACAGACTTTGCTGCTTACAGCACTCATGCTCATCA	373	100%	0	0	0	0
Query 242	AGTTGAGAGGAAATCGGTGCGAGTGTGACACATCCGAGAGATGAGTATTCCTTCGAGT	301	100%	0	0	0	0
Db 374	AGTTGAGAGGAAATCGGTGCGAGTGTGACACATCCGAGAGATGAGTATTCCTTCGAGT	433	100%	0	0	0	0
Query 302	GCCCTACCGCGGAGAACTTGTGCTCGTTCGCTGGGCTGCTGCTGCGAAGCGCAGAA	361	100%	0	0	0	0
Db 434	GCCCTACCGCGGAGAACTTGTGCTCGTTCGCTGGGCTGCTGCTGCGAAGCGCAGAA	493	100%	0	0	0	0
Query 362	TGCCATCCGCTGTCAGCTGCGTGAACGTGTGGGTGCTGTGAGGAGTCTGCAGTAAGC	421	100%	0	0	0	0
Db 494	TGCCATCCGCTGTCAGCTGCGTGAACGTGTGGGTGCTGTGAGGAGTCTGCAGTAAGC	553	100%	0	0	0	0
Query 422	TCTATGAGACCCGCTGTACACACCCAGCATGTTCTGCGCCGCGGAGGAGGCAAGCAAGG	481	100%	0	0	0	0
Db 554	TCTATGAGACCCGCTGTACACACCCAGCATGTTCTGCGCCGCGGAGGAGGCAAGCAAGG	613	100%	0	0	0	0
Query 482	ACTCCTGCAAGGATCTCTGTGGGGGCGCCGTGATCTGCAACGGGATCTTGCAGGGCCCTTG	541	100%	0	0	0	0
Db 614	ACTCCTGCAAGGATCTCTGTGGGGGCGCCGTGATCTGCAACGGGATCTTGCAGGGCCCTTG	673	100%	0	0	0	0
Query 542	TGCTTTTGGGAAAAGCCCGGTGTGGCCCACTTGGCGTCCAGGTGTCTACACCAACTCT	601	100%	0	0	0	0
Db 674	TGCTTTTGGGAAAAGCCCGGTGTGGCCCAAGTTGGGGTCCAGGTGTCTACACCAACTCT	733	100%	0	0	0	0
Query 602	GCAAAATTCATGATGATAGAGAAAACCGTTCAGNCCAGTTAACTCTGGGACATGGGAA	661	100%	0	0	0	0
Db 734	GCAAAATTCATGATGATAGAGAAAACCGTTCAGNCCAGTTAACTCTGGGACATGGGAA	793	100%	0	0	0	0

AC X41114.
 DT 17-JUN-1999 (first entry)
 DE Human secreted protein 5' EST SEQ ID NO:58.
 KW Human: secreted protein; EST: expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW differentiation; haematopoiesis regulation; tissue growth regulation;
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
 OS Homo sapiens.
 PN W0906548-A2.
 PD 11-FEB-1999.
 PF 31-JUL-1998; IB1222.
 PR 01-AUG-1997; US-905135.
 PA (GIST) GENSET.
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;
 DR WPI: 99-153778/13.
 DR P-PSDB: Y12281.
 PT New nucleic acids encoding human secreted proteins - obtained from
 PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
 PT kidney, lung, umbilical cord, placenta and colon tissue
 PS Claim 1, Page 205; 824pp; English.
 CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
 CC secreted proteins, and encode the proteins given in Y12261 to Y12514,
 CC respectively. The proteins given represent the signal peptide and an
 CC N-terminal fragment of a secreted protein. The nucleic acid sequences
 CC can be used for producing secreted human gene products. They can also
 CC be used to develop products for diagnosis and therapy. The proteins
 CC obtained may have cytokine activity, cell proliferation/differentiation
 CC activity, haematopoiesis regulating activity, tissue growth regulating
 CC activity, reproductive hormone regulating activity, chemotactic/
 CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
 CC ligand activity, anti-inflammatory activity, tumour inhibition activity
 CC or other activities. The products can be used in forensic, gene therapy
 CC and chromosome mapping procedures. The sequences can also be used for
 CC obtaining corresponding promoter sequences. The nucleic acids encoding
 CC the signal peptide can be used for directing extracellular secretion of
 CC a polypeptide or the insertion of a polypeptide into a membrane, or
 CC importing a polypeptide into a cell.
 SO Sequence 402 BP; 80 A; 125 C; 121 G; 75 T;

Query Match 28.4%; Score 331.4; DB 1; Length 402;
 Best Local Similarity 96.2%; Pred. No. 5.5e-74;
 Matches 354; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 Oy 2 CGACGCTGCGAGCGGCATGTCATGGAAGCAATTTGCTCGGGCTCTGG 61
 Db CGACGCTGCGAGCGGCATGTCATGGAAGCAATTTGCTCGGGCTCTGG 88
 Oy 62 TGCATCCCGCAGTGGTGTCTGTCAGCCGCACACTGTTTCCAGAA-----CTCCT 109
 Db TGCATCCCGCAGTGGTGTCTGTCAGCCGCACACTGTTTCCAGAA-----CTCCT 148
 Oy 89 TGCATCCCGCAGTGGTGTCTGTCAGCCGCACACTGTTTCCAGAAAGTAGTCKAGACTCT 148
 Db TGCATCCCGCAGTGGTGTCTGTCAGCCGCACACTGTTTCCAGAAAGTAGTCKAGACTCT 148
 Oy 110 ACACCATGCGGGCTGGCCCTGCACAGTCTTGAGCGGACCAAGAGCGAGGAGCAAGATGG 169
 Db ACACCATGCGGGCTGGCCCTGCACAGTCTTGAGCGGACCAAGAGCGAGGAGCAAGATGG 208
 Oy 149 ACACCATGCGGGCTGGCCCTGCACAGTCTTGAGCGGACCAAGAGCGAGGAGCAAGATGG 208
 Db ACACCATGCGGGCTGGCCCTGCACAGTCTTGAGCGGACCAAGAGCGAGGAGCAAGATGG 208
 Oy 170 TGGAGGCGACGCTCTCCGTAAGCGCAACAGATACAAACAGACTTGTGCTGCTAAGCAG 229
 Db TGGAGGCGACGCTCTCCGTAAGCGCAACAGATACAAACAGACTTGTGCTGCTAAGCAG 268
 Oy 209 TGGAGGCGACGCTCTCCGTAAGCGCAACAGATACAAACAGACTTGTGCTGCTAAGCAG 268
 Db TGGAGGCGACGCTCTCCGTAAGCGCAACAGATACAAACAGACTTGTGCTGCTAAGCAG 268
 Oy 230 TCATCTCATCAAGTTGGAGCAATCCGTTCCGAGTCTGACACCAATCCGGAGCATCAGCA 289
 Db TCATCTCATCAAGTTGGAGCAATCCGTTCCGAGTCTGACACCAATCCGGAGCATCAGCA 328
 Oy 269 TCATCTCATCAAGTTGGAGCAATCCGTTCCGAGTCTGACACCAATCCGGAGCATCAGCA 328
 Db TCATCTCATCAAGTTGGAGCAATCCGTTCCGAGTCTGACACCAATCCGGAGCATCAGCA 328
 Oy 290 TTGCTTCGAGTGCCTACCGCGGGGAACCTTTGGCTGCTGTTCTGGCTGGGCTGCTGG 349
 Db TTGCTTCGAGTGCCTACCGCGGGGAACCTTTGGCTGCTGTTCTGGCTGGGCTGCTGG 388
 Oy 329 TTGCTTCGAGTGCCTACCGCGGGGAACCTTTGGCTGCTGTTCTGGCTGGGCTGCTGG 388
 Db TTGCTTCGAGTGCCTACCGCGGGGAACCTTTGGCTGCTGTTCTGGCTGGGCTGCTGG 388
 Oy 350 CGAAGGCG 357
 Db CGAAGGCG 396

RESULT 12
 ID V58646 standard; cDNA; 1459 BP.
 AC V58646;
 DT 08-DEC-1998 (first entry)
 DE Prostate tumour specific gene clone DE6.
 KW Prostate tumour specific gene; human; prostate cancer; detection;
 KW therapy; SS.
 KW Homo sapiens.
 PN W09837418-A2.
 PD 27-AUG-1998.
 PF 25-FEB-1998; U03690.
 PR 09-FEB-1998; US-904809.
 PR 25-FEB-1997; US-806596.
 PR 01-AUG-1997; US-904809.
 PA (CORI-) CORIXA CORP.
 PI Dillon DC, Xu J;
 DR WPI: 98-480805/41.
 PT Novel human prostate specific tumour protein and fragments - useful
 PT for detecting and treating prostate cancers
 PS Claim 1, Page 114; 141pp; English.
 CC This sequence represents a human prostate tumour specific gene, and can
 CC be used in the method of the invention. The method is for detecting
 CC prostate cancer comprises contacting a biological sample with an agent
 CC able to bind an immunogenic portion of a prostate protein (such as
 CC encoded by this sequence). An antibody which binds to an immunogenic
 CC portion of the prostate protein, and the method can be used to detect,
 CC monitor progression of, or treat prostate cancers. The antibody may
 CC also be conjugated to a therapeutic agent for use in therapy of prostate
 CC cancers.
 SO Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Query Match 25.1%; Score 292.8; DB 1; Length 1459;
 Best Local Similarity 80.4%; Pred. No. 3.4e-64;
 Matches 390; Conservative 0; Mismatches 12; Indels 83; Gaps 1;
 Oy 92 ACTGTTCCAGAACTCTTACACCATCGGGCTGGGCTGCACAGTCTTGGAGCCGACCAAG 151
 Db ACTGTTCCAGAACTCTTACACCATCGGGCTGGGCTGCACAGTCTTGGAGCCGACCAAG 151
 Oy 25 AGTAGTGCAGAGCTCCATACACCATCGGGCTGGGCTGCACAGTCTTGGAGCCGACCAAG 84
 Db AGTAGTGCAGAGCTCCATACACCATCGGGCTGGGCTGCACAGTCTTGGAGCCGACCAAG 84
 Oy 152 AGCCAGGAGGACATAGTGTGAGGCGCACCCTCTCCGTACGGCACCAGAGTACCAAGAC 211
 Db AGCCAGGAGGACATAGTGTGAGGCGCACCCTCTCCGTACGGCACCAGAGTACCAAGAC 211
 Oy 85 AGCCAGGAGGACATAGTGTGAGGCGCACCCTCTCCGTACGGCACCAGAGTACCAAGAC 144
 Db AGCCAGGAGGACATAGTGTGAGGCGCACCCTCTCCGTACGGCACCAGAGTACCAAGAC 144
 Oy 212 TCTTCTGCTTACAGACCTCATGCTCATCAAGTTGAGAGTAATCGGTCCGAGTCTGACA 271
 Db TCTTCTGCTTACAGACCTCATGCTCATCAAGTTGAGAGTAATCGGTCCGAGTCTGACA 271
 Oy 145 CTTTCTGCTTACAGACCTCATGCTCATCAAGTTGAGAGTAATCGGTCCGAGTCTGACA 204
 Db CTTTCTGCTTACAGACCTCATGCTCATCAAGTTGAGAGTAATCGGTCCGAGTCTGACA 204
 Oy 272 CCATCCGGAGGATACAGATTGCTTCCGAGTCCCTTACCGCGGGGAACCTTGGCTCGTNT 331
 Db CCATCCGGAGGATACAGATTGCTTCCGAGTCCCTTACCGCGGGGAACCTTGGCTCGTNT 331
 Oy 205 CCATCCGGAGGATACAGATTGCTTCCGAGTCCCTTACCGCGGGGAACCTTGGCTCGTNT 264
 Db CCATCCGGAGGATACAGATTGCTTCCGAGTCCCTTACCGCGGGGAACCTTGGCTCGTNT 264
 Oy 332 CTGGCTGGGGTCTGCTGGCGAAC-----
 Db CTGGCTGGGGTCTGCTGGCGAAC-----
 Oy 265 CTGGCTGGGGTCTGCTGGCGAACGGTGAAGCTACAGGGTGTGTCTGCTCTTCAAGGA 324
 Db CTGGCTGGGGTCTGCTGGCGAACGGTGAAGCTACAGGGTGTGTCTGCTCTTCAAGGA 324
 Oy 355 -----GGCAGAAATGCTTAC 368
 Db -----GGCAGAAATGCTTAC 384
 Oy 325 GGTCTCTGCGCCAGTCCGGGGGCTGACCCAGAGACTTGTGCTCCAGGACAGAAATGCTTAC 384
 Db GGTCTCTGCGCCAGTCCGGGGGCTGACCCAGAGACTTGTGCTCCAGGACAGAAATGCTTAC 384
 Oy 369 CGTGTGACAGTGTGAGCAAGTGTGGTGTCTGAGAGANGTCTGACAGTAACTCTTATGA 428
 Db CGTGTGACAGTGTGAGCAAGTGTGGTGTCTGAGAGANGTCTGACAGTAACTCTTATGA 444
 Oy 385 CGTGTGACAGTGTGAGCAAGTGTGGTGTCTGAGAGANGTCTGACAGTAACTCTTATGA 444
 Db CGTGTGACAGTGTGAGCAAGTGTGGTGTCTGAGAGANGTCTGACAGTAACTCTTATGA 444
 Oy 429 CCCGCTGACACCCAGCAATGTTCTGCGCGCGGGAGGAGCAAGCAAGAAAGACTCTG 488
 Db CCCGCTGACACCCAGCAATGTTCTGCGCGCGGGAGGAGCAAGCAAGAAAGACTCTG 504
 Oy 445 CCCGCTGACACCCAGCAATGTTCTGCGCGCGGGAGGAGCAAGCAAGAAAGACTCTG 504
 Db CCCGCTGACACCCAGCAATGTTCTGCGCGCGGGAGGAGCAAGCAAGAAAGACTCTG 504
 Oy 489 CAACG 493
 Db CAACG 493

Db 505 CAACG 509

RESULT 13

ID V61251 standard; cDNA: 1459 BP.
 AC V61251;
 DT 06-JAN-1999 (first entry)
 DE cDNA sequence of prostate tumour clone P703 splice variant D6.
 KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
 OS Homo sapiens.
 PN MO9837053-A2.
 PD 27-AUG-1998.
 PF 25-FEB-1998; 003492.
 PR 09-FEB-1998; US-020956.
 PR 25-FEB-1997; US-806099.
 PR 01-AUG-1997; US-904804.
 PA (CORI-) CORIXA CORP.
 PI Dillon DC, Xu J;
 DR WPI: 98-609886/51.
 PT Polypeptides comprising immunogenic portions of prostate proteins - used in a vaccine for the treatment of prostate cancer
 PS Claim 3: Page 106, 130pp; English.
 CC The present sequence is a new DNA which encodes an immunogenic portion of a prostate tumour protein. The encoded immunogen, or the DNA itself, can be used as a vaccine for the treatment of prostate cancer. The DNA was identified by analysis of a subtracted cDNA library obtained by CC subtracting a prostate tumour cDNA expression library with a normal CC tissue cDNA library.
 SQ Sequence 1459 BP: 427 A; 328 C; 406 G; 295 T;

Query Match

Best Local Similarity 25.1%; Score 292.8; DB 1; Length 1459;
 Matches 390; Conservative 0; Mismatches 12; Indels 83; Gaps 1;

QY 92 ACTGTTCCAGAACTCTACACCATCGGGGCTGGGCTGCACAGCTTTGAGGCCGACCAAG 151
 Db 25 AGTAGATGACAGACTCTCTACACCATCGGGGCTGGGCTGCACAGCTTTGAGGCCGACCAAG 84
 QY 152 AGCCAGGAGCCAGATGTTGGAGGCCAGCTCTCCGTAAGGACCCAGATACACAGAC 211
 Db 85 AGCCAGGAGCCAGATGTTGGAGGCCAGCTCTCCGTAAGGACCCAGATACACAGAC 144
 QY 212 TCTTGCTCGCTAAGCACTCATCTCATCAAGTTGAGCAATCCGTTCCGAGTGTGACA 271
 Db 145 CTTGCTCGCTAAGCACTCATCTCATCAAGTTGAGCAATCCGTTCCGAGTGTGACA 204
 QY 272 CCATCCGAGCATACGATTTGCTTCGAGTGCCTTACCGGGGGAATCTTGGCTGTTNT 331
 Db 205 CCATCCGAGCATACGATTTGCTTCGAGTGCCTTACCGGGGGAATCTTGGCTGTTNT 264
 QY 332 CTGGCTGGGGTCTGCTGGCGCAAC----- 355
 Db 265 CTGGCTGGGGTCTGCTGGCGCAACGTTGAGCTACGGGTGTGTGCTCCCTCTCAAGGA 324
 QY 355 -----GGCAGAAATGCCCTAC 368
 Db 325 GGTCCTCTGCGCCAGTGGCGGGGGCTGACCCAGAGCTCTGGCTCCAGGCAAGATGCCCTAC 384
 QY 369 CGTGTGACATGGGTGAACGTGTGGTGTGTGAGAGANGTGTGACAGTCTATGA 428
 Db 385 CGGTGTGACATGGGTGAACGTGTGGTGTGTGAGAGANGTGTGACAGTCTATGA 444
 QY 429 CCGGCTGTACACCCAGCATGTTCTGCGCGCGGAGGAGGACAGACAGAGACTCTG 488
 Db 445 CCGGCTGTACACCCAGCATGTTCTGCGCGCGGAGGAGGACAGACAGAGACTCTG 504
 QY 489 CAACG 493
 Db 505 CAACG 509

RESULT 14

ID V58522 standard; cDNA: 234 BP.
 AC V58522;
 DT 08-DEC-1998 (first entry)
 DE Prostate tumour specific gene clone P20.
 KW Prostate tumour specific gene; human; prostate cancer; detection; therapy; ss.
 OS Homo sapiens.
 PN MO9837418-A2.
 PD 27-AUG-1998.
 PF 25-FEB-1998; 003690.
 PR 09-FEB-1998; US-904809.
 PR 25-FEB-1997; US-806596.
 PR 01-AUG-1997; US-904809.
 PA (CORI-) CORIXA CORP.
 PI Dillon DC, Xu J;
 DR WPI: 98-480805/41.
 PT Novel human prostate specific tumour protein and fragments - useful for detecting and treating prostate cancers
 PS Claim 1: Page 56; 141pp; English.
 CC This sequence represents a human prostate tumour specific gene, and can be used in the method of the invention. The method is for detecting prostate cancer comprises contacting a biological sample with an agent able to bind an immunogenic portion of a prostate protein (such as CC encoded by this sequence). An antibody which binds to an immunogenic portion of the prostate protein, and the method can be used to detect, CC monitor progression of, or treat prostate cancers. The antibody may CC also be conjugated to a therapeutic agent for use in therapy of prostate CC cancers.
 SQ Sequence 234 BP: 43 A; 68 C; 68 G; 55 T;

Query Match

Best Local Similarity 19.6%; Score 228.8; DB 1; Length 234;
 Matches 230; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 203 ACACAGACTCTTGTGCTGCTAAGCACTCATCTCATCAAGTTGAGCAATCCGTTCCG 262
 Db 1 ACACAGACTCTTGTGCTGCTAAGCACTCATCTCATCAAGTTGAGCAATCCGTTCCG 60
 QY 263 AGTGTACACCAATCCGAGCATACGATTTGCTTCGAGTGCCTTACCGGGGGAATCTT 322
 Db 61 AGTGTACACCAATCCGAGCATACGATTTGCTTCGAGTGCCTTACCGGGGGAATCTT 120
 QY 323 GCCTCTTCTGCTGCTGGGCTGCTGCGGACGAGCAATGCCCTACCGGCTGCTGCG 382
 Db 121 GCCTCTTCTGCTGCTGGGCTGCTGCGGACGAGCAATGCCCTACCGGCTGCTGCG 180
 QY 383 TGAACGTGTGCGGTGTGCTGAGAGANGTGTGCAAGTCTATGACCCGCTGT 436
 Db 181 TGAACGTGTGCGGTGTGCTGAGAGANGTGTGCAAGTCTATGACCCGCTGT 234

RESULT 15

ID V61168 standard; cDNA: 234 BP.
 AC V61168;
 DT 06-JAN-1999 (first entry)
 DE cDNA sequence of prostate tumour clone P20.
 KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
 OS Homo sapiens.
 PN MO9837093-A2.
 PD 27-AUG-1998.
 PF 25-FEB-1998; 003492.
 PR 09-FEB-1998; US-020956.
 PR 25-FEB-1997; US-806099.
 PR 01-AUG-1997; US-904804.
 PA (CORI-) CORIXA CORP.
 PI Dillon DC, Xu J;
 DR WPI: 98-609886/51.
 PT Polypeptides comprising immunogenic portions of prostate proteins - used in a vaccine for the treatment of prostate cancer

PS Claim 3; Page 53-54; 130bp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 234 BP; 43 A; 68 C; 68 G; 55 T;

Query Match 19.68; Score 228.8; DB 1; Length 234;
Best Local Similarity 98.38; Pred. No. 1.9e-48;
Matches 230; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 203 ACAACAGACTCTTGGCTGCTAAGCAGACCTCATCAAGTTGAGAGAAATCCGTGCCG 262
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 ACAACAGACCCCTTCTGCTGCTAAGCAGACCTCATCAAGTTGAGAGAAATCCGTGCCG 60

QY 263 AGTCTGACACCATCCGAGCATCAGATGCTTGCAGTGCCTACCGCGGGAATCTTT 322
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 61 AGTCTGACACCATCCGAGCATCAGATGCTTGCAGTGCCTACCGCGGGAATCTTT 120

QY 323 GCCTCGTTCTGGCTGGGCTGCTGCTGCGAAGCGCAGATGCTTACCGTGTGCACTGCC 382
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 121 GCCTCGTTCTGGCTGGGCTGCTGCTGCGAAGCGCAGATGCTTACCGTGTGCACTGCC 180

QY 383 TGAACGTTGGTGGTGTGCTGAGAGANGTCTGACAGTCTATGACCCGCTGT 436
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 181 TGAACGTTGGTGGTGTGCTGAGAGANGTCTGACAGTCTATGACCCGCTGT 234
```

Search completed: September 25, 1999, 09:57:15
Job time: 8086 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 12:05:23 ; Search time 1811.29 seconds

(without alignments)
1270.888 million cell updates/sec

Title: US-09-030-606-175

Perfect score: 1167
Sequence: 1 GCGCAGCCCTGCGAGCGGC.....NTAAGAGAGCGCAAAAAA 1167

Scoring table: IDENTITY_NOC

Searched: 2546578 seqs; 98626752 residues

Database :

EST:
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: em_est20:*
21: em_est21:*
22: em_est22:*
23: em_est23:*
24: em_est24:*
25: em_est25:*
26: em_est26:*
27: em_est27:*
28: em_est28:*
29: em_est29:*
30: em_est30:*
31: em_est31:*
32: em_est32:*
33: em_est33:*
34: em_est34:*
35: em_est35:*
36: em_est36:*
37: em_est37:*
38: em_est38:*
39: em_est39:*
40: em_est40:*
41: em_est41:*
42: em_est42:*
43: em_est43:*
44: em_est44:*
45: em_est45:*
46: em_est46:*
47: em_est47:*
48: em_est48:*
49: em_est49:*
50: em_est50:*
51: em_est51:*
52: em_est52:*
53: em_est53:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	555.6	47.6	777	48	AI557281	AI557281 PT2.1.15-
2	403.2	34.6	569	50	AI686689	AI686689 t035g11.x
3	393	33.7	415	35	AA551449	AA551449 n155e05.s
4	348	29.8	404	34	AA503963	AA503963 nh39a01.s
5	286.6	24.6	576	34	AA533140	AA533140 n146h06.s
6	221	18.9	732	48	AI557025	AI557025 PT2.1.10-
7	217.2	18.6	259	35	AA565161	AA565161 nk52c07.s
8	210.8	18.1	229	35	AA552779	AA552779 nk57a10.s
9	203	17.4	218	50	AI674521	AI674521 wc39g02.x
10	201.2	17.2	241	36	AA603977	AA603977 no47a09.s
11	174	14.9	188	50	AI675815	AI675815 wb97b12.x
12	164	14.1	216	32	AA336074	AA336074 EST40886
13	161.8	13.9	191	35	AA595489	AA595489 no34h04.s
14	132	11.3	457	45	AI385433	AI385433 ml85b05.y
15	118.2	10.1	498	26	W73168	W73168 zd55e11.r1
16	116	9.9	505	45	AI391329	AI391329 mb71h07.y
17	114	9.8	496	39	AA664127	AA664127 vn42e07.r
18	112.6	9.6	507	35	AA583052	AA583052 n080e09.s
19	106.4	9.1	517	31	AA293027	AA293027 zt54a12.r
20	100.2	8.6	530	43	AA411252	AA411252 zc33b03.r
21	99.2	8.5	759	45	AI326340	AI326340 ml61e08.x
22	99	8.4	321	31	AA302930	AA302930 EST113097
23	98.4	8.4	686	46	AI415008	AI415008 mb71h07.x
24	98	8.4	586	26	W73140	W73140 zd55e11.s1
25	97.4	8.3	626	36	AA643312	AA643312 nr59b08.s
26	96	8.2	333	49	AI636241	AI636241 tz93c03.x
27	92.6	7.9	356	39	AA838788	AA838788 a187g08.s
28	91.4	7.8	212	28	C16642	C16642 C16642 Clon
29	90.2	7.7	420	43	AI226226	AI226226 ue8f06.y
30	90	7.7	585	35	C23111	C23111 C23111 Japa
31	90	7.7	504	49	AI620091	AI620091 ty48i10.x
32	89.8	7.7	480	39	AA846771	AA846771 a141f01.s
33	89.8	7.7	367	41	AI002101	AI002101 ct38d03.s
34	89.6	7.7	234	33	AA452459	AA452459 zx29g09.r
35	89	7.6	400	31	AA293231	AA293231 zt26g09.r
36	89	7.6	539	34	AA477689	AA477689 zu44a12.r
37	87.8	7.5	467	28	AA073833	AA073833 mj99h09.r
38	87.8	7.5	460	38	AA791893	AA791893 vs54h07.r
39	87.4	7.5	519	33	AA401397	AA401397 zu68h01.s
40	87.4	7.5	583	40	AA921373	AA921373 ak56d12.s
41	87.4	7.5	493	45	AI324874	AI324874 ml85b09.s
42	87.2	7.5	541	39	AA844955	AA844955 ak61d09.s
43	87.2	7.5	324	43	AI177474	AI177474 EST221106
44	87.2	7.5	599	43	AI237604	AI237604 EST34166
45	87.2	7.5	457	46	AA988593	AA988593 UI-R-CO-1

ALIGNMENTS

RESULT 1
AI557281 777 bp mRNA
LOCUS PT2.1.15_G12.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
DEFINITION AI557281
ACCESSION AI557281
NID 94489644
VERSION AI557281.1 GI:4489644
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS 1 (bases 1 to 777)
Huang, G.M., Ng, W., Farakas, J., Chen, L., Liang, H.A., Gordon, D., Jun Yu, J., and Hood, L.
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138767.

Contact: Guyang Mathew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanqgm@yahoo.com.

FEATURES
source
1..777
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="960H1: 6; 6p21.31-6p22.1; 21q"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional cDNA library was constructed using lambda ZP II kit (Stratagene). mRNA was extracted from a frozen prostate tumor tissue (Mayo Clinics)."
BASE COUNT 162 a 205 c 227 g 157 t 26 others
ORIGIN

Query Match 47.6%; Score 555.6; DB 48; Length 777;
Best Local Similarity 91.8%; Pred. No. 1.3e-128;
Matches 637; Conservative 0; Mismatches 46; Indels 11; Gaps 5;

2 CGAGAGCCCTGGAGGCGGCACTGGTATGGAAGCAATGTTCTGCTGGGGCTCTTG 61
21 CGAGAGCCCTGGAGGCGGCACTGGTATGGAAGCAATGTTCTGCTGGGGCTCTTG 80
62 TGCATCCGAGTGGTGTCTGACGCGCACACTGTTTCCAGAACTCTTACACCATGGGC 121
81 TGCATCCGAGTGGTGTCTGACGCGCACACTGTTTCCAGAACTCTTACACCATGGGC 140
122 TGGGCTGACAGTCTTGAGGCGGACCAAGAGCCAGATGTTGAGAGCCAGCC 181
141 TGGGCTGACAGTCTTGAGGCGGACCAAGAGCCAGATGTTGAGAGCCAGCC 200
182 TCTCCGTAGCGGACCCAGAGTACAGAGATCTTGTCTGCTAGAGACCTCATCTCA 241
201 TCTCCGTAGCGGACCCAGAGTACAGAGATCTTGTCTGCTAGAGACCTCATCTCA 260
242 AGTTGACGATCCGTGTCTGACGACATCCGAGACATCAGCATTTGCTTGCAGT 301
261 AGTTGACGATCCGTGTCTGACGACATCCGAGACATCAGCATTTGCTTGCAGT 320
302 GCCCTACCGGGGGAAGTCTGCTGTCGTCGGGTGCTGCGGCAAGCCAGAA 361
321 GCCCTACCGGGGGAAGTCTGCTGTCGTCGGGTGCTGCGGCAAGCCAGAA 380
362 TGCCTACCGTGTCTGACGAGTGTGAGTGTGAGGAGTGTGAGGAGTGTGAGGAG 421
381 TG-CTACCGTGTCTGACGAGTGTGAGTGTGAGGAGTGTGAGGAGTGTGAGGAG 439
422 TCTATGACCGGCTGTACACACCCAGCATGTTCTGCGCGGCGGAGGAGCAAGAGG 481
440 TCTATGACCGGCTGTACACACCCAGCATGTTCTGCGCGGCGGAGGAGCAAGAGG 499
482 ACTCCCTGCAACGGTGTACTGGGGGGCCCTGATCTGCAACGGGTACTTGCAGGCTTG 541
500 ACTCCCTGCAACGGTGTACTGGGGGGCCCTGATCTGCAACGGGTACTTGCAGGCTTG 558

542 TGCCTTGGAAAGCCCGTGGCAACTTGGCGTGGCAGGTGTACACCACTCT 601
559 TGTCTTTGGAAAGCCCGTGG--CAAGTTGGCTGTGAGTGTACACAC-----CT 611
602 GCAAATTCAGTGTGATAGAGAAACCGTCCAGNCCAGTTACTGTGGGAGTGGAA 661
612 CTGAATTAATCAGTGTGATAGAGAAACCGTCCAG--CAGTTAACTTNGGGNACTGNA 669
662 CCCATGAATTCAGTGTGATAGAGAAACCGTCCAGTCCGCGGA 695
670 NCCATTAAATTCAGTGTGATAGAGAAACCGTCCGCGGA 703

RESULT 2
A1686689/c 569 bp mRNA EST 27-MAY-1999
LOCUS tu35g11.x1 NCI_CGAP_P128 Homo sapiens cDNA clone IMAGE:2253092 3'
DEFINITION similar to TR:092046 092046 PREPROTRYPsin PRECURSOR ;, mRNA sequence.
ACCESSION A1686689
KEYWORDS NID 94897983
VERSION A1686689.1 GI:4897983
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 569)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
TITLE Unpublished (1997)
JOURNAL On Jun 5, 1998 this sequence version replaced gi:3189584.
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www-bio.lnl.gov/bbrp/image/image.html

Seq primer: -400p from gibco
High quality sequence stop: 444.
Location/Qualifiers
1..569
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2253092"
/clone_lib="NCI_CGAP_P128"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_P128 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 125 a 150 c 170 g 124 t
ORIGIN

Query Match 34.6%; Score 403.2; DB 50; Length 569;
Best Local Similarity 98.8%; Pred. No. 1.1e-90;
Matches 405; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 151 GAGCAGGAGGACGAGATGATGAGGAGCCAGCCCTCTCCGTACGCGACCCAGAGTACACAGA 210
 |||||||
 Db 569 GAGCAGGAGGACGAGATGATGAGGAGCCAGCCCTCTCCGTACGCGACCCAGAGTACACAGA 510
 |||||||
 QY 211 CTCTTCTCTGCTTAACGACCTTATGCTATCAAGTTGACGAGATCCGTCTCCGAGTCTGAC 270
 |||||||
 Db 509 CCTCTGCTGCTTAACGACCTTATGCTATCAAGTTGACGAGATCCGTCTCCGAGTCTGAC 450
 |||||||
 QY 271 ACCATCCGAGGACATCAGATGCTTCCGAGTCCCTACCGGGGGGAGACTCTTGGCTGCTN 330
 |||||||
 Db 449 ACCATCCGAGGACATCAGATGCTTCCGAGTCCCTACCGGGGGGAGACTCTTGGCTGCTN 390
 |||||||
 QY 331 TCTGCTGAGGCTCTGCTGCGCAACGCGAGATGCTTACCGCTGCTGCTGCTGCTGCTGCTG 390
 |||||||
 Db 389 TCTGCTGAGGCTCTGCTGCGCAACGCGAGATGCTTACCGCTGCTGCTGCTGCTGCTGCTG 330
 |||||||
 QY 391 TCGGTGCTGCTGAGANGTCTGCAATGAGCTTATGACCCGCTGTACACCCAGCATG 450
 |||||||
 Db 329 TCGGTGCTGCTGAGANGTCTGCAATGAGCTTATGACCCGCTGTACACCCAGCATG 270
 |||||||
 QY 451 TTCTGCGCCGCGGAGGAGGAGACAGAGGAGCTCTGCAACGCTGCTGCTGCTGCTGCTGCTG 510
 |||||||
 Db 269 TTCTGCGCCGCGGAGGAGGAGACAGAGGAGCTCTGCAACGCTGCTGCTGCTGCTGCTGCTG 210
 |||||||
 QY 511 CTGATCTGCAACGCGGCTACTTGCAGGCGCTTGTCTTTCGGAAGAGCCCC 560
 |||||||
 Db 209 CTGATCTGCAACGCGGCTACTTGCAGGCGCTTGTCTTTCGGAAGAGCCCC 160
 |||||||

RESULT 3
 LOCUS AA551449 415 bp mRNA EST 05-SEP-1997
 DEFINITION n355805.s1 NCI_CGAP_P9 Homo sapiens cDNA clone IMAGE:96416
 similar to SW:K1KA_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR
 ; mRNA sequence.
 ACCESSION AA551449
 NID 92321701
 VERSION AA551449.1 GI:2321701
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 415)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 9, 1995 this sequence version replaced gi:802243.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuagui,
 M.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: David B. Kitzman, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www.bio.lnu.gov/bdrip/image/image.html

Insert Length: 640 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 412.
 Location/Qualifiers

FEATURES
 source 1..415
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="96416"
 /clone_id="NCI_CGAP_P9"
 /sex="male"

/tissue_type="normal prostatic epithelial cells"
 /lab_host="DH10B"
 /note="Organ: prostate; Vector: PAMPI0; mRNA made from
 normal prostatic epithelial cells, cDNA made by oligo-dt
 priming. Non-directionally cloned. Size selected on
 agarose gel, average insert size 600 bp. Library made by
 D. Kitzman, NIH."
 BASE COUNT 78 a 125 c 129 g 83 t
 ORIGIN

Query Match 33.7%; Score 393; DB 35; Length 415;
 Best local Similarity 98.1%; Pred. No. 3.3e-88;
 Matches 407; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 137 TTGAGGCGGACCAAGAGCCAGGAGCCAGATGATGAGGAGCCAGCTTCTCCGTACGGCACC 196
 |||||||
 Db 2 TCGAGCGCCGACCAAGAGCCAGGAGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 61
 |||||||
 QY 197 CAGAGTACACAGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 256
 |||||||
 Db 62 CAGAGTACACAGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 121
 |||||||
 QY 257 TGTCCAGTGTGACACCATCCGAGCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 316
 |||||||
 Db 122 TGTCCAGTGTGACACCATCCGAGCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 181
 |||||||
 QY 317 ACTCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 376
 |||||||
 Db 182 ACTCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
 |||||||
 QY 377 ACTGCTGACGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 436
 |||||||
 Db 241 AGTCTGACGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
 |||||||
 QY 437 ACCACCCAGCATGTTCTGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 496
 |||||||
 Db 301 ACCACCCAGCATGTTCTGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
 |||||||
 QY 497 ACTCTGCGGCG 551
 |||||||
 Db 361 ACTCTGCGGCG 415
 |||||||

RESULT 4
 LOCUS AA503963/c 404 bp mRNA EST 20-AUG-1997
 DEFINITION nh39401.s1 NCI_CGAP_P5 Homo sapiens cDNA clone IMAGE:954696
 similar to SW:K1KA_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR
 ; contains MSRL.b2 MSRL repetitive element ; mRNA sequence.
 ACCESSION AA503963
 NID 92238930
 VERSION AA503963.1 GI:2238930
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 404)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 8, 1995 this sequence version replaced gi:801235.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuagui,
 M.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: David B. Kitzman, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1244 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 144.

FEATURES

```

source
1. .404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:954696"
/clone_id="NCI_CGAP_Pt5"
/sex="male"
/tissue_type="prostate"
/lab_host="DH10B"
/note="Vector: pAMP10; mRNA made from normal prostatic
epithelial cells, cDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel
average insert size 600 bp."
BASE COUNT
64 a 91 c 159 g 90 t
ORIGIN

```

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP)
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
 On Sep 12, 1996 this sequence version replaced gi:1406821.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuquila,
M.D., Michael R. Emmert-Stuck, M.D., Ph.D.
cDNA library Preparation: David B. Krizman, Ph.D.
cDNA library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/bdnp/image/image.html

FEATURES

```

Insert Length: 885      Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 256.
      Location/Qualifiers
          1. 576

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1 IMAGE:995579"
/clone_1lib="NCI_CGAP_P19"
/sex="male"
/tissue_type="normal prostatic epithelial cells"
/lab_host="DH10B"
/notes="Organ: prostate; Vector: pAMP10; mRNA made from normal prostatic epithelial cells, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Library made by D. Krizman, NIH."
BASE COUNT      169 a      156 c      139 g      102 t      10 others
ORIGIN

```


RESULT	9
A1674521/c	
LOCUS	
DEFINITION	A1674521 218 bp mRNA Homo sapiens CDNA IMAGE:2321042 3
	wc39g02.x1 NCI_CGAP_P1r78 Homo sapiens CDNA IMAGE:2321042 3

[illegible]


```

RESULT 10
AA603977/c 241 bp mRNA EST 28-OCT-1997
LOCUS nc047a09.s1 NCI_CGAP_Pr23 Homo sapiens CDNA clone IMAGE:1103800 3'
DEFINITION similar to contains MSRI.D3 MSRI repetitive element ;, mRNA
sequence.
ACCESSION AA603977
NID 92444547
VERSION AA603977.1 GI:2444547
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 241)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1395285.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrip/image/image.html

Insert Length: 2084 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 231.
Location/Qualifiers
1. 241
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1103800"
/clone_1ib="NCI_CGAP_Pr23"
/sex="male, pooled"
/tissue="prostate tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: prostate; Vector: Bluescript SK-. Site_1:
EcORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pooled prostate tumors. 5' adaptor sequence: 5'
GAATTCGGCAGAG 3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."
BASE COUNT 46 a 37 c 78 g 80 t
ORIGIN

Query Match 17.2%; Score 201.2; DB 36; Length 241;
Best Local Similarity 96.2%; Pred. No. 1.6e-40;
Matches 203; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 957 GAGGTCCAGAGCCCAACCCCTGTCCTCCAGACCCAGAGTCCAGGCTCCAGCCCTCC 1016
|||||
DB 241 GAGGTCCAGAGCCCAACCCCTGTCCTCCAGACCCAGAGTCCAGGCTCCAGCCCTCC 182
|||||

QY 1017 TCCTCTAGACCCAGGCTCCCAATGACCTGANTTCCCTGTACACAGTCCCTTTGT 1076
|||||
DB 181 TCCTCTAGACCCAGGCTCCCAATGACCTGANTTCCCTGTAAACAGTCCCTTTGT 122
|||||

QY 1077 GGCAGTGTGACCCAGCTTACCACTGGTTTTCATTTTGTCCCTTCCCTAGATCC 1136
|||||
DB 121 GGCAGTGTGACCCAGCTTACCACTGGTTTTCATTTTGTCCCTTCCCTAGATCC 62
|||||

QY 1137 AGAATTAAGTNTAAGAGAGCGCAAAAAA 1167
|||||
DB 61 AGAATTAAGTNTAAGAGAGCGCAAAAAA 31

```

```

RESULT 11
A1675815/c 188 bp mRNA EST 19-MAY-1999
LOCUS wb97b12.x1 NCI_CGAP_Pr28 Homo sapiens CDNA clone IMAGE:2313599 3',
DEFINITION mRNA sequence.
ACCESSION A1675815
NID 94876295
VERSION A1675815.1 GI:4876295
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 188)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Dec 20, 1995 this sequence version replaced gi:1133540.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrip/image/image.html

Seq primer: -40UP from Gibco.
Location/Qualifiers
1. 188
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2313599"
/clone_1ib="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonoids
985608-986759, 1101192-1101959, and 1217928-1220615)."
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 43 a 31 c 61 g 53 t
ORIGIN

Query Match 14.9%; Score 174; DB 50; Length 188;
Best Local Similarity 97.2%; Pred. No. 8.9e-34;
Matches 174; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 989 ACCCAGAGTNCAGAGTCCAGCCCTCTCTCTCAGACCCAGGAGGTCATATGCCACTAG 1048
|||||
DB 188 ACCCAGAGTNCAGAGTCCAGCCCTCTCTCTCAGACCCAGGAGGTCATATGCCACTAG 129
|||||

QY 1049 ANTNTCCCTGTACACAGTGGCCCCCTTGCGCANGTTGACCAACCTTACAGTTGTTT 1108
|||||
DB 128 ACNTCCCTGTACACAGTGGCCCCCTTGCGCANGTTGACCAACCTTACAGTTGTTT 69
|||||

QY 1109 TCATTTTGTCTCCCTTCCCTAGATCCAGAAATTAAGTTTAAGAGAGCGCAAAAAA 1167
|||||
DB 68 TCATTTTGTCTCCCTTCCCTAGATCCAGAAATTAAGTTTAAGAGAGCGCAAAAAA 10

```

RESULT 12
AA336074 216 bp mRNA EST 21-APR-1997
LOCUS AA336074
DEFINITION EST0886 Endometrial tumor Homo sapiens cDNA 5' end similar to similar to kallikrein family, mRNA sequence.
ACCESSION AA336074
NID 91988560
VERSION AA336074.1 GI:1988560
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 216)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Feldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man,Wal,C., Clifton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaro,S.M., Merrick,J.M., Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,O.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Fertile,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L., Wei,X.F., Wang,J., Xu,C., Yu,G.L., Rubin,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 96026280
COMMENT On Apr 14, 1993 this sequence version replaced gi:692773.

TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018689056
Fax: 3018689423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES
source
1. 216
/organism="Homo sapiens"
/db_xref="ATCC (Inhost):138014"
/db_xref="taxon:9606"
/clone_lib="Endometrial tumor"
/sex="female"
/dev_stage="adult"
/note="Organ: endometrium; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 45 a 65 c 63 g 38 t 5 others
ORIGIN

Query Match 14.1%; Score 164; DB 32; Length 216;
Best Local Similarity 97.0%; Pred. No. 2.9e-31;
Matches 164; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 CGCAGCCCTGGAGGCGGACCTGTCATGTAAGAAACGAATGTTCTCTGGGGGCTCTGG 61
DB 48 CGCAGCCCTGGAGGCGGACCTGTCATGTAAGAAACGAATGTTCTCTGGGGGCTCTGG 107
QY 62 TGCATCCGAGTGGGTGCTGTACGCCGACACTGTTCCAGAACTCTACACCATGGGCG 121

DB 108 TGCATCCGAGTGGGTGCTGTACGCCGACACTGTTCCAGAACTCTACACCATGGGCG 167
QY 122 TGGCCCTGCACAGTCTTGTAGGCGCCGACGAGCCGAGGAGCCAGTGTGT 170
DB 168 TGGCCCTGCACAGTCTTGTAGGCGCCGACGAGCCGAGGAGCCAGTGTGT 216
RESULT 13
AA595489 191 bp mRNA EST 18-SEP-1997
LOCUS AA595489/C
DEFINITION n034h04.s1 NCI_CGAP_P123 Homo sapiens cDNA clone IMAGE:1102615 3', mRNA sequence.
ACCESSION AA595489
NID 92410839
VERSION AA595489.1 GI:2410839
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 191)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
JOURNAL
TITLE
AUTHORS
COMMENT On Apr 14, 1993 this sequence version replaced gi:692625.

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNI at: www.bio.lnlnl.gov/db/rp/image/image.html

FEATURES
source
1. 191
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1102615"
/clone_lib="NCI_CGAP_P123"
/sex="male, pooled"
/tissue_type="prostate tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: prostate; Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Pooled prostate tumors. 5' adaptor sequence: 5' GAATTCGACGAG 3' 3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3' Average insert size: 1.2 kb."
BASE COUNT 43 a 28 c 57 g 63 t
ORIGIN

Query Match 13.9%; Score 161.8; DB 35; Length 191;
Best Local Similarity 95.9%; Pred. No. 9.8e-31;
Matches 163; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 998 TMCAGTCCGAGGCGGCGGCTCTCTCCAGACGAGGCGGTCCATGCGACCTAGANTTCCT 1057
DB 191 TCCAGGTACCAAGCCCTCTCTCTCTAGACCGAGGCGGTCCATGCGACCTAGANTTCCT 132
QY 1058 GTACACAGTGGCCCTTGTGCGANTGTGACCAACCTTACAGAGTGTGTTTCATTTT 1117
DB 131 GTACACAGTGGCCCTTGTGCGANTGTGACCAACCTTACAGAGTGTGTTTCATTTT 72
QY 1118 GTCCCTTCCCTAGATCCAGAAATTAAGTNTAAGAGAGGCAAAAAA 1167
DB 71 GTCCCTTCCCTAGATCCAGAAATTAAGTNTAAGAGAGGCAAAAAA 22


```
/clone_lib="Soares_fetal_heart_NbH19W"  
/sex="unknown"  
/dev_stage="19 weeks"  
/lab_host="DH10B (ampicillin resistant)"  
/note="Organ: heart; Vector: pT7/3D (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer (5'  
TGTTCACCAATCTGAGTGGAGCGCGGCGATCTTTTCTTTTCTTTT 3'),  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT7/3 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by  
M.Fatima Bonaldo. This library was constructed from the  
same fetus as the fetal lung library, Soares fetal lung  
NbH19W."
```

BASE COUNT 116 a 150 c 128 g 104 t

ORIGIN

```
Query Match                    10.1%; Score 118.2; DB 26; Length 498;  
Best Local Similarity    61.0%; Pred. No. 1e-19;  
Matches 227; Conservative    0; Mismatches 136; Indels    9; Gaps    2;  
  
QY    275 TCCGGAGCATCAGCATTCGTCGAGTGCCTACCGCGGGGAACCTCTGCTCGTCTG    334  
      |||    |||    |||    |||    |||    |||    |||    |||    |||    |||  
DB    25 TCAGACCCATCAAGCTCTCTCATTCCTCTGCTGCGGACAAAGTCTGTGTCTG    84  
  
QY    335 GCTGGGGTCTGCTGGGAGGCGGAGATG-----CCTACCGTGTGCACTGGGTGAACG    388  
      |||||    |||    |||    |||    |||    |||    |||    |||    |||    |||  
DB    85 GCTGGGGGACACCAAGAGCCCGGAGTCACTTCCCTAAGTCTCCAGTGTGAATA    144  
  
QY    389 TGTGGTGTGTCTGAGAGAGTCTGACATAGCTCTATGACCCGCTGTACACCCAGCA    448  
      |||    |||    |||    |||    |||    |||    |||    |||    |||    |||  
DB    145 TCAGCGTGTCTAAGTACAAAAGGTGAGAGTGTTCACCGAGACAGATAGATACACCA    204  
  
QY    449 TGTTCGCGCGCGGCGGAGGCGAAGACAGACTCTGCAACGGTACTGTGGGGGGC    508  
      |||||    |||    |||    |||    |||    |||    |||    |||    |||    |||  
DB    205 TGTCTGCGCGCGGTGACA---AAGCAGGTAGAGACTCTGCCAGGGTGTGTTGGGGGC    261  
  
QY    509 CCCGATCTGCAAGGGTACTGTGAGGGGCTTGTGCTTTCGGAAGCCCGGTGGCC    568  
      |||    |||    |||    |||    |||    |||    |||    |||    |||    |||  
DB    262 CTGTGTGTCTGCAAGGCTCTGCGAGGAGCTCTGTCTGTGGGAGATTACCTGTGCCC    321  
  
QY    569 AACTGGCGTGCAGAGTGTCTACACCAACCTCTGCAATTCACCTGAGTGAAGAGAAA    628  
      |||    |||    |||    |||    |||    |||    |||    |||    |||    |||  
DB    322 GGCCCAACAGACCGGGGTGTACAGCAACCTGTGCAAGTTCAACCAAGTGAATCCAGSAA    381  
  
QY    629 CGGTCCAGNCCA    640  
      |||    |||    |||    |||    |||    |||    |||    |||    |||    |||  
DB    382 CCATCCAGGCCA    393
```

Search completed: September 25, 1999, 12:05:28
Job time: 8132 sec


```

|||||
84  rgsrllseerllealaserGlnCysProthrAlaGlnSerCysLeu 100
334 GTTCTGGCGGGGCTGCTGGGGAACGATGCTGATTCACATCCATC 383
101 ValSerGlyTrpIleuLeuAlaAsnAspAlaValIleAlaIleGlnSe 117
384 CCAGACTGTGGAGGCTGGAGTGTGAGAAAGCTTCCCAACCTGGCAGG 433
117 r***ThrValGlyGlyTrpIleuCysGlnCysLeuSerGlnProtrpGln 134
434 GTTGTACATTTGGGCACTTCAGTGCAGAGAGAGCTCTCTGATCCTC 483
134 LysThrIleSerAlaThrSerSerAlaArgThrSerCysCysIleLeu 150
484 ACTGGGCTCACTACTGCTCACTGCATCACCAGGAACTG 525
151 ThrGlyCysSerLeuLeuThrAlaSerProGlyThrLeu 164

seq_name: A_Geneseq_36:W71873

```

```

seq_documentation_block:
ID W71873 standard; Protein: 164 AA.
AC W71873.
DE 06-JAN-1999 (first entry)
DE Protein encoded by prostate tumour clone P703 splice variant DE14.
KW Prostate; cancer; tumour; vaccine; immunogen; clone.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 118 /note= "undefined residue"
PN MO9837093-A2.
PD 27-AUG-1998.
PR 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806089.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Example 3; Page 109; 130pp; English.
CC The present sequence is an immunogenic portion of a prostate tumour
CC protein. The immunogen, or the DNA encoding it, can be used as a
CC vaccine for the treatment of prostate cancer. The immunogen was
CC isolated from a prostate tumour cDNA library obtained by subtracting
CC a prostate tumour cDNA expression library with a normal tissue cDNA
CC library.
SQ Sequence 164 AA:

```

```

alignment_scores:
Quality: 854.00 Length: 164
Ratio: 5.239 Gaps: 0
Percent Similarity: 99.390 Percent Identity: 99.390

```

alignment_block:

US-09-030-606-177 x W71873 ..

Align seg 1/1 to: W71873 from: 1 to: 164

```

34 ATGGAAGCAATTTGTTCTGCTGGGCGCTCTGTCATCGCAGTGGGT 83
|||||
1 MetGlnAsnGlnIleuPheCysSerGlyValLeuValHisProGlnTrpVal 17
84 GGTGTACGCGGCACTGTTTCAGAACTCTACACCATGGGGCTGGCC 133
|||||
17 IleuSerAlaAlaHisCysPheGlnAsnSerTyrThrIleGlyLeuGlyL 34
134 TGCACAGTCTTTAGAGCCGACCAAGAGCCAGGAGCCAGATGGTGGAGCC 183
|||||
34 euhISerLeuGlnAlaAspGlnIleProGlySerGlnMetValGlnAla 50

```

```

184 AGCCTCCGCTACGAGCCAGACCAAGTACAAGACACCTGTGCTGCTACGA 233
|||||
51 SerLeuSerValArgHisProGlnIleTrpAsnArgProLeuAlaAsnAs 67
234 CCTCATGCTCATCATCAAGTTGAGCAATCCGTGTCGAGTGTGACCATCC 283
67 PleuMetLeuIleLysLeuAspIleuSerValSerGlnSerAspThrIleA 84
284 GAGCATACAGATTCGTTGCAATGCCCTACCGCGGGAACTTCTCCCTC 333
84 rgsrllseerllealaserGlnCysProthrAlaGlnSerCysLeu 100
334 GTTCTGGCGGGGCTGCTGGGGAACGATGCTGATTCACATCCATC 383
101 ValSerGlyTrpIleuLeuAlaAsnAspAlaValIleAlaIleGlnSe 117
384 CCAGACTGTGGAGGCTGGAGTGTGAGAAAGCTTCCCAACCTGGCAGG 433
117 r***ThrValGlyGlyTrpIleuCysGlnCysLeuSerGlnProtrpGln 134
434 GTTGTACATTTGGGCACTTCAGTGCAGAGAGAGCTCTCTGATCCTC 483
134 LysThrIleSerAlaThrSerSerAlaArgThrSerCysCysIleLeu 150
484 ACTGGGCTCACTACTGCTCACTGCATCACCAGGAACTG 525
151 ThrGlyCysSerLeuLeuThrAlaSerProGlyThrLeu 164

seq_name: A_Geneseq_36:W59129

```

```

seq_documentation_block:
ID W59129 standard; Protein: 232 AA.
AC W59129.
DE 11-SEP-1998 (first entry)
DE Homo sapiens Tub Interactor (hTl-1) protein.
DE serine protease; tub interactor; treatment; obesity; cachexia;
DE anorexia nervosa; diabetes; cell cycle progression; apoptosis;
DE neurodegenerative disease; Alzheimer's disease; drug screening;
DE Parkinson's disease; Huntington's chorea; detection; diagnosis;
DE amyotrophic lateral sclerosis; spinocerebellar degeneration.
KW Homo sapiens.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Region 42 /note= "undefined amino acid"
PN MO9812302-A1.
PD 26-MAR-1998.
PR 05-SEP-1997; U15627.
PR 21-JUL-1997; US-897340.
PR 17-SEP-1996; US-715032.
PA (MILL-) MILLENNIUM PHARM INC.
PI Errada PR, Gimeno CJ;
DR WPI: 98-217246/19.
DR N-PSDB: V11835.
PT Tub interactor genes - used to develop products for the treatment
PT of obesity, cachexia, anorexia nervosa or related disorders e.g.
PT diabetes
PS Claim 28; Fig 1; 120pp; English.
CC The sequence is that encoding the Tub Interactor protein (hTl-1)
CC which is a putative serine protease. TI genes function
CC in biochemical pathways involved in weight control and
CC related disorders. The products can be used for treating
CC weight disorders, e.g. obesity, cachexia or anorexia nervosa,
CC or a related disorder such as diabetes. The products can
CC also be used to modulate cell cycle progression and apoptosis.
CC They can be used for treating neurodegenerative diseases
CC which are characterised by apoptosis, including Alzheimer's
CC disease, Parkinson's disease, Huntington's chorea, amyotrophic
CC lateral sclerosis or spinocerebellar degenerations. The
CC products can also be used for detection, diagnosis and
CC drug screening.
SQ Sequence 232 AA:

```

alignment_scores:

Quality: 621.00 Length: 154
Ratio: 4.669 Gaps: 3
Percent Similarity: 86.364 Percent Identity: 81.169

alignment_block:

US-09-030-606-177 x W59129 ..

Align seg 1/1 to: W59129 from: 1 to: 232

```

1 GCGCACTGCGACCCCTGGCAGCGCGCACTGTCATGAGAAACGAATTGTT 50
5 AIAHISSEGLNPRGTPGlnAlaAlaLeuValMetGluAsnGlnLeuPh 21
51 CTGCTGGGGCGTCTGTCATCCGAGAGGGTGTCTGACGGCAACT 100
21 ecyserylvalleuvalhisproglintrpvalleuserialaahisc 38
101 GTTTC.....CAGAACTCTACACCATCGCGCTGGCTGCAC 138
38 yspheglnlys**ValGlnserSeryThrIleGlyLeuGlyLeuHis 54
139 AGCTTGAAGCCGACCAAGCCAGGAGCCAGATGTGGAGCCAGCCT 188
55 SerLeuGlnAlaAspGlnGlnProGlySerGlnMetValGlnAlaSerle 71
189 CTGCGTAGCGCACCGAGTACAGACACCTTGCTGCTACGACCTCA 238
71 userValAlaHisProGlnTrpAsnArgProLeuAlaAsnAspLeuM 88
239 TGGTCATCAAGTTGGACGATCCGTGTCGAGTGTACACATCCGAGAC 288
88 eLeuIleIleIleuAspGlnuserValSerGlnSerAspThrIleArgSer 104
289 ATCAGCAATGCTTTCAGTGCCTACCGCGGGGAACTTGGCTGTTTC 338
105 IleserIleAlaSerGlnCysProThrAlaGlyAsnSerCysLeuValSe 121
339 TGGCTGGGGGTCTGCTGGCGAGCATGCTGTG..... 369
121 rGlyTrpIleLeuAlaAsnGlyArgMetProThrValLeuGlnCysV 138
370 .....ATTGCCATCCAGTCCGAGCTGGGAGGCTGGAGTGTGAGAG 414
138 alaAsnValSerValAlaSerGlnGlnVal.....CysSerlys 150
415 CTTCGCCAACCC 426
151 LeuTyraAspPro 154

```

seq_name: A_Geneseq_36:W60592

seq_documentation_block:

ID W60592 standard; Protein: 248 AA.
AC W60592;
DT 07-SEP-1998 (first entry)
DE Human prostate-specific kallikrein (HPSK) protein.
KW Prostate-specific kallikrein; HPSK; prostate carcinoma; human;
benign prostate hyperplasia; diagnosis; drug screening; PSK.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 113 /label= unknown
FT /note= "encoded by NTC"
FT Misc_difference 128 /label= unknown
FT /note= "unknown
FT Misc_difference 132 /note= "encoded by AGN"
FT /label= unknown
FT /note= "encoded by GNT"
PN W09820117-A1.
PD 14-MAY-1998.

PF 31-OCT-1997; U20051.
PR 05-NOV-1996; US-744026.
PA (INCY-) INCYTE PHARM INC.
PI Bandman O, Goll SR.
DR WPI; 98-286933/25.
DR N-PSDB; V37495.
PT New isolated prostate-specific kallikrein - used to develop products
for diagnosis and treatment of, e.g. prostate carcinoma or benign
hyperplasia
PS Claim 1. Fig 1A-C, 68pp; English.
CC This represents a human prostate-specific kallikrein (HPSK). A host cell
containing an expression vector comprising the HPSK nucleic acid sequence
can be used to produce the protein recombinantly. The HPSK products can
be used for the diagnosis of conditions or diseases associated with
expression of HPSK such as prostate carcinoma and benign prostate
hyperplasia. Agonists and antagonists which specifically bind to HPSK and
modulate its activity can be used for the preparation of treatment of
such conditions or diseases. The products can also be used for detection
and drug screening, especially for the detection of prostate-specific
kallikrein (PSK).
SO Sequence 248 AA:

alignment_scores:

Quality: 610.00 Length: 149
Ratio: 4.692 Gaps: 2
Percent Similarity: 87.248 Percent Identity: 80.537

alignment_block:

US-09-030-606-177 x W60592 ..

Align seg 1/1 to: W60592 from: 1 to: 248

```

4 CACTGCGACGCCCTGGCAGCGCGCACTGTCATGAGAAACGAATTGTTG 53
34 HisSerGlnProTrpGlnAlaAlaLeuValMetGlnGlnGlnLeuPh 50
54 CTGCGGCGTCTGGTGCATCCGAGTGGGCTGTGTCAGCGCCACACTTT 103
50 sSerGlyValLeuValHisProGlnTrpValLeuSerIleAlaHisCysP 67
104 TCCAGATCTCTACACATCCGCGGCTGGGCTGTGCACAGTGTGAGGCCAC 153
67 heGlnAsnSerTyrThrIleGlyLeuGlyLeuHisSerLeuGlnAlaAsp 83
154 CAAGAGCCAGGAGCCAGATGTGGAGGCCAGCTTCGTCGAGGACCC 203
84 GlnGlnProGlySerGlnMetValGlnAlaSerLeuSerValArgHisPr 100
204 AGAGTACACAGACCCCTGCTGCTGCTACGACCTCATGCTCATCAAGTTG 253
100 oGlnTrpAsnArgProLeuAlaAsnAspLeuMet**IleLysLeuA 117
254 ACGAATCCGTGTCGAGTGTACACATCCGAGGACGAGCATGAGCATGCTG 303
117 spGlnuserValSerGlnuserAspAsnIleArg***IleSerIle***Ser 133
304 CAGTGCCTTACCGCGGGGAACTCTGCTGCTTCTGCTGGGGTGTGCT 353
134 GlnCysProThrAlaGlyAsnPhcysLeuValSerGlyTrpGlyLeu 150
354 GCGCAACGATGCTGTG.....ATTGCCATCC 379
150 uAlaAsnGlyArgMetProThrValLeuGlnCysValAsnValSerVal 167
380 AGTCCAGACTGTGGAGAGCTGGAGAGTGTGAGAGCTTCCCAACCC 426
167 alSerGlnGlnVal.....CysSerlysLeuTyraAspPro 178

```

seq_name: A_Geneseq_36:W69388

seq_documentation_block:

ID W69388 standard; Protein: 205 AA.


```

284 GGAGCATCAGCATGCTTCGACGTGCCCTACCCGCGGGAACCTGCTC 333
|||||
84 rgsrllleerlleaialaserlncysprrhrhlaaglsansercysleu 100
|||||
334 GTTTCGGCTGGGGTCTGCTGGCGAACGATGCTGTG..... 369
|||||
101 valserglytrpilyleuualaaanglyargmetprrhrvalleunh 117
|||||
370 .....ATGCCATCCAGTCCAGACAGCTGGAGCGCTGGAGTGTG 409
|||||
117 scysvalasnvalservalvalserlsu**Val.....CysS 130
|||||
410 AGAGCTTTCCCAACCC 426
|||||
130 erlylsleuylrnsppro 135
|||||
seq_name: A_Geneseq_36:W69387

```

```

seq_documentation_block:
ID W69387 standard; Protein; 159 AA.
AC W69387;
DE 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE1 protein.
KW Prostate tumour specific gene; human; prostate cancer; detection;
  therapy.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 103 /note= "unspecified amino acid"
FT Misc_difference 105 /note= "unspecified amino acid"
FT Misc_difference 105 /note= "unspecified amino acid"
PN W09837418-A2.
PD 27-AUG-1998.
PE 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
PI WPI: 98-460805/41.
DR N-PSDB; V58644.
PT Novel human prostate specific tumour protein and fragments - useful
  for detecting and treating prostate cancers
PS Example 1: Page 112-113; 14pp; English.
CC This sequence is encoded by a human prostate tumour specific gene, and
  can be used in the method of the invention. The method is for detecting
  prostate cancer comprising contacting a biological sample with an agent
  able to bind an immunogenic portion of a prostate protein (such as
  this protein sequence). An antibody which binds to an immunogenic
  portion of the prostate protein, and the method can be used to detect,
  monitor progression of, or treat prostate cancers. The antibody may
  also be conjugated to a therapeutic agent for use in therapy of prostate
  cancers.
CC Sequence 159 AA:

```

```

alignment_scores:
  Quality: 328.00      Length: 93
  Ratio: 4.316        Gaps: 2
Percent Similarity: 81.720 Percent Identity: 74.194

```

alignment_block:
US-09-030-606-177 x W69387 ..

Align seg 1/1 to: W69387 from: 1 to: 159

```

172 ATGGTGAGGCGCAGCTTCGACGCGACCCAGATCAAGACACCTT 221
|||||
1 MetvalGluAlaSerleuSerValArgHisProGluTyrAsnArgProle 17
|||||
222 GCTCGCTAACGACCTCATGCTCAAGTTGAGAGATCCGTCGAGT 271
|||||

```

```

17 uleuAlaAsnAspLeuMetLeuIleLysLeuAspGluSerValSerGluS 34
272 CTGACACCATTCGCGAGCATACGATTCGTCGACGCCCTACCGGGG 321
|||||
34 eraSphrllleayserlleerlleaialaserlncysprrhrhlaagly 50
|||||
322 AACTCTGGCTGCTGCTGCTGGCGTGTGCTGGCGAACGATGCTGTG... 369
|||||
51 AsnserCysleuValserlytrpilyleuualaaanglyargmetprr 67
|||||
370 .....ATGCCATCCAGTCCAGACAGCTGGAGCGCTGGAGTGTG 397
|||||
67 orhrvalleunGlyncysvalasnvalservalvalserlsu**Val..... 82
|||||
398 GCTGGAGCTGTGAGAGCTTTCCCAACCC 426
|||||
83 .....CysSerlylsleuylrnsppro 89
|||||
seq_name: A_Geneseq_36:W71871

```

```

seq_documentation_block:
ID W71871 standard; Protein; 159 AA.
AC W71871;
DE 06-JAN-1999 (first entry)
DE Protein encoded by prostate tumour clone P703 splice variant DE1.
KW Prostate; cancer; tumour; vaccine; immunogen; clone.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 103 /note= "undefined residue"
FT Misc_difference 105 /note= "undefined residue"
FT Misc_difference 105 /note= "undefined residue"
PN W09837093-A2.
PD 27-AUG-1998.
PE 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
PI WPI: 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
  used in a vaccine for the treatment of prostate cancer
PS Example 3: Page 105; 130pp; English.
CC The present sequence is an immunogenic portion of a prostate tumour
  protein. The immunogen, or the DNA encoding it, can be used as a
  vaccine for the treatment of prostate cancer. The immunogen was
  isolated from a prostate tumour cDNA library obtained by subtracting
  a prostate tumour cDNA expression library with a normal tissue cDNA
  library.
CC Sequence 159 AA:

```

```

alignment_scores:
  Quality: 328.00      Length: 93
  Ratio: 4.316        Gaps: 2
Percent Similarity: 81.720 Percent Identity: 74.194

```

alignment_block:
US-09-030-606-177 x W71871 ..

Align seg 1/1 to: W71871 from: 1 to: 159

```

172 ATGGTGAGGCGCAGCTTCGACGCGACCCAGATCAAGACACCTT 221
|||||
1 MetvalGluAlaSerleuSerValArgHisProGluTyrAsnArgProle 17
|||||
222 GCTCGCTAACGACCTCATGCTCAAGTTGAGAGATCCGTCGAGT 271
|||||
17 uleuAlaAsnAspLeuMetLeuIleLysLeuAspGluSerValSerGluS 34
|||||
272 CTGACACCATTCGCGAGCATACGATTCGTCGACGCCCTACCGGGG 321
|||||

```

34 *erasphriliteargserilileserilealasergincysprothralagly* 50
322 *AACTCTTGCTCGTTCCTGCTGGGGGTCTGCTGGCAACAGTGGTGG*. 368
|||||
|||||
51 *Asnsercysleuvalsercyltrpclipyleuualaasnlyargmetpr* 67
370ATGGCATCCAGTCCAGACTGTGGAG 397
67 *othrvalleugincysvalasvalservalvalserlunlval*.... 82
398 *GCTGGAGAGTGGAGAGCTTCCCAACC* 426
83cyssterlystleuylrasppro 89
|||||
|||||

seq_name: A_Geneseq_36:Y12281

```

seq_documentation_block:
ID Y12281 standard; protein; 66 AA.
AC Y12281;
DT 17-JUN-1999 (first entry)
DE Human 5' EST secreted protein SEQ ID NO:312.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokine; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.
OS Homo sapiens.
PN W09906548-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IBI222.
PR 01-AUG-1997; US-905135.
PA (GSEST ) GENSEST
PI Ducleert A, Dumaz Malne Edwards J, Lacroix B;
DR WPI: 99-153778/13.
DR N-PDSB; X41114.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
PS Claim 27: Page 661: 824pp: English
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulation
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokine activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 66 AA;

```

222 GCTCGCTAACGACGCTCATCTGCATCAATGGACGAATCCGTGCCAGT 27
17 uleuallaasaspIemelleIleIetyIenaspIuIserValserGus 34
272 CTGACACCAATCCGGAGCAATCAGCAATGGCTGGCAGTGGCCACGGCGGG 32
34 erspPhrIleAgserrIleesrIleIleIasGrcInGcyrIrothIaIagly 50
322 AACTCTTGCCCTGTTTCTGCGCTGGGGGCTGCTGCGGCAAC 360
51 AsnIserCysIeuuValserIlyrIpIglyIeuuIleuIlaas 63

seq_name: A_Geneseq_36:W10694

seq_documentation block:	
ID	W10694 standard; Protein: 260 AA.
AC	W10694;
DT	08-APR-1997 (first entry)
DE	Human recombinant neurotrophin-4, used for antibody production.
KW	Antibody; Alzheimer's disease; neurodegenerative; epitope; epilepsy.
CC	recombinant.
OS	Homo sapiens.
FT	Key
FT	peptide
FT	Location/Qualifiers
FT	95..215
FT	/note="claimed fragment of the protein, used for antibody production"
PN	J08245700-A.
PD	24-SEP-1996.
PF	14-MAR-1995; 083154.
PR	14-MAR-1995; JP-083154.
PA	(IGAK)- IGAKU SEIBUTSUGAKU KENKYUSHO KK.
DR	WPI; 96-482259/48.
DR	N-PSDB; T48519.
PT	Antibody against neurotrophin-4 used in the treatment of Alzheimer's disease and epilepsy
PS	Claim 1, Page 7; 9pp. Japanese.
CC	W10694 is a recombinantly produced human neurotrophin protein. The main invention relates to an antibody against the neurotrophin protein, in particular a claimed fragment of the neurotrophin protein (see W10694 and features table). The antibody can be used to treat neurodegenerative disorders such as Alzheimer's disease and epilepsy
SQ	Sequence 260 AA;

```

107 ocstYrnsnansSerAsnProGluasphHisSerHisaspIleMetLeu 124
245 TCAAGTTGGACGAATCCGTTCGAGTCGACATCCAGCATCCGGAGCATCAGC 294
124 leaYglueGlnasSerHisAlasnleuGlyaspLysvalLysProvalGln 140
295 ATTCGCTTCGAGTGCCTACCGCGGGAGAACTTCCTGCTTTCGGCTG 344
141 LeuAlasnleuGlyProLysvalGlyGlnLysCysIleIleSerGlyTr 157
345 GGGTCTGCTGGCG.....AACGATG 364
157 pGlyThrValThrSerProGlnGluAsnPheProAsnThrLeuAsnCysA 174
365 CTGTGATTCGATCCGATCCGAGCTGTGGAGCGTGGAGTGTGAGAG 414
174 laGluValLysIleThrSerGlnAsn.....LysCysGlnuArg 186
415 CTT.....TCCCAACCTTCGCGAGGTTGTATCATTTCGGCAACTTCGAG 458
187 AlaTrpProGlyLysIleThrGlnGlyMetValCysAlaGlySerSerAs 203
459 TGCAGAGCAGCTCTGC 474
203 nGlyAlaAspThrCys 208
seq_name: A_Geneseq_36:W12393

```

```

seq_documentation_block:
ID W12393 standard: Protein: 260 AA.
AC W12393;
DT 15-MAY-1997 (first entry)
DE Mouse neuropilin protein.
KW Mouse; neuropilin; hippocampus; lambda gt10; primer; PCR; amplification;
KW polymerase chain reaction; serine protease domain; nerve growth factor;
KW NGF; insect cell; virus; expression vector; transfection;
KW cerebral disease.
OS Mus musculus.
PN J08311099-A.
PD 26-NOV-1996.
PF 13-MAR-1996: 056367.
PR 14-MAR-1995: JP-054584.
PA (SHIO/) SHIOZAKA S.
DR WPI: 97-061812/06.
DR N-PSDB: T63251.
PT Nucleic acid encoding neuropilin - for producing neuropilin, useful
PT for diagnosis and treatment of cerebral disease
PS Claim 1; Page 6-7; 9pp; Japanese.
CC This is the amino acid sequence of a novel mouse protein designated
CC neuropilin. The encoding gene was isolated from a mouse hippocampal
CC cDNA library in lambda gt10 using a cloned, amplified fragment of
CC the gene (clone B41, T63254). This fragment was amplified using
CC primers T63252-3. The primers were synthesised based on the serine
CC protease domain of nerve growth factor (NGF)-gamma. The screen isolated
CC 6 positive clones, of which clone NP5 contained the longest insert
CC (this sequence). The protein has a molecular weight of around 26 kD.
CC It has 43% homology with EGF-BP 41 & with NGF-gamma; 39% with NGF-alpha;
CC 38% with trypsin and 18% with tPA. The protein can be used for clinical
CC diagnosis and treatment of cerebral diseases.
SQ Sequence 260 AA;

```

```

alignment_scores:
Quality: 296.00 Length: 172
Ratio: 2.446 Gaps: 5
Percent Similarity: 70.349 Percent Identity: 35.465

```

```

alignment_block:
US-09-030-606-177 x W12393

```

```

Align seg 1/1 to: W12393 from: 1 to: 260

```

```

4 CACTTCGACCCCTGGCAGCGGCGACTGTCATGAAACGAATGTCTCG 53

```

```

|||||..... |||..... :: ..... |||
42 HisSerGlnProTrpGlnAlaAlaLeuPheGlnGlyGluArgLeuIleCys 58
54 CTCGGGCTCTCTGTCATCCGAGTGGTGTCTTCACGCCACACGCTT 103
58 sGlyValLeuValGlyAspArgTrpValLeuThrAlaAlaHisCysL 75
104 TCCGAACCTCTACACCATCGGGCTGGGCTTCGACAGCTTCGAGCCGAC 153
75 yLysGlnLysTrpSerValArgLeuGlyAspHisSerLeuGlnSerArg 91
154 CAAGAGCCAGGAGCAGATGTCGAGGCCACCTTCCTCGTACGCAACC 203
92 AspGlnPro...GluGlnGluIleGlnValAlaGlnSerIleGlnHisSpr 107
204 AGAGTACACAGACCTCTG.....CTCGTACAGACCTCTATGCTCA 244
107 ocstYrnsnansSerAsnProGluasphHisSerHisaspIleMetLeu 124
245 TCAAGTTGGACGAATCCGTTCGAGTCGACATCCAGCATCCGGAGCATCAGC 294
124 leaYglueGlnasSerHisAlasnleuGlyaspLysvalLysProvalGln 140
295 ATTCGCTTCGAGTGCCTACCGCGGGAGAACTTCCTGCTTTCGGCTG 344
141 LeuAlasnleuGlyProLysvalGlyGlnLysCysIleIleSerGlyTr 157
345 GGGTCTGCTGGCG.....AACGATG 364
157 pGlyThrValThrSerProGlnGluAsnPheProAsnThrLeuAsnCysA 174
365 CTGTGATTCGATCCGATCCGAGCTGTGGAGCGTGGAGTGTGAGAG 414
174 laGluValLysIleThrSerGlnAsn.....LysCysGlnuArg 186
415 CTT.....TCCCAACCTTCGCGAGGTTGTATCATTTCGGCAACTTCGAG 458
187 AlaTrpProGlyLysIleThrGlnGlyMetValCysAlaGlySerSerAs 203
459 TGCAGAGCAGCTCTGC 474
203 nGlyAlaAspThrCys 208
seq_name: A_Geneseq_36:W94493

```

```

seq_documentation_block:
ID W94493 standard: Protein: 268 AA.
AC W94493;
DT 23-APR-1999 (first entry)
DE Human kallikrein.
KW Human; kallikrein; keratinocyte; HKALL; skin disorder; cancer; eczema;
KW psoriasis; scleroderma; adenocarcinoma; leukemia; melanoma.
OS Homo sapiens.
PN M09842849-A1.
PD 01-OCT-1998.
PF 25-MAR-1998: U05939.
PR 26-MAR-1997: US-824874.
PA (INCY-) INCYTE PHARM INC.
PA Hillman JL, Lal P;
PI WPI: 99-070073/06.
DR N-PSDB: X16295.
PT Human kallikrein polypeptide, HKALL - useful e.g. to treat skin
PT disorders e.g. eczema, psoriasis and screen for antagonists useful
PT to treat skin disorders and cancers
PS Claim 1; Page 41-42; 61pp; English.
CC The present sequence represents human kallikrein, designated HKALL.
CC HKALL (or fragments) can be administered therapeutically to increase
CC proteolysis and subsequent skin scaling to treat/prevent skin disorders,
CC e.g. eczema, psoriasis and scleroderma. HKALL has chemical and
CC structural homology with human stratum corneum chymotryptic enzyme
CC (thought to be involved in the proteolysis of intercellular cohesive
CC structures necessary for desquamation, the process by which outer layers
CC of skin are eliminated), and its expression in cDNA libraries was

```

CC associated with tumour-associated tissues and skin cells. It can be used
CC to screen for antagonists and agonists, and to generate antibodies.
CC HKAL1 antagonists can be used to suppress excessive proteolysis and
CC subsequent skin cell scaling, so may be administered to treat skin
CC disorders. They may also be used to suppress excessive cell
CC proliferation, so can be administered to treat/prevent cancer, e.g.
CC adenocarcinoma, leukemia and melanoma. Antibodies specific for HKAL1
CC may be used directly as antagonists, or indirectly as a targeting or
CC delivery mechanism for bringing pharmaceutical agents to HKAL1-expressing
CC cells. They are also useful to diagnose conditions/diseases characterised
CC by HKAL1 expression and to monitor therapeutic interventions. The
CC polynucleotide encoding HKAL1, or complementary sequences, can be used to
CC produce hybridisation probes, useful to detect polynucleotides encoding
CC HKAL1, e.g. to diagnose diseases relating to polypeptide expression
CC (e.g. cancers of the bladder, prostate) or monitor HKAL1 regulation
CC during therapeutic intervention. Polynucleotides encoding HKAL1 are
CC useful to produce antisense sequences for therapeutic administration to
CC modulate/prevent HKAL1 expression e.g. to treat/prevent skin disorders
CC or cancer as above.
SQ Sequence 268 AA;

alignment_scores: Quality: 292.50 Length: 142
 Ratio: 2.985 Gaps: 3
Percent Similarity: 69.014 Percent Identity: 42.254

alignment_block:
US-09-030-606-177 x W94493 ..

Align seg 1/1 to: W94493 from: 1 to: 268

```
19 CAGCGCGCAGTGGTCATGAGAA...AACGAATGTTCTGTCGGGCGTCT 65
   |||||.....:||||:||||:|||||
56 GlnAlaIaIaLeuLeuLeuArgProAsnGlnLeuTyrCysGlyAlaValLe 72
   |||||.....:||||:||||:|||||
66 GGTGCATCCGACAGTGGTGTGTCAGCCGACACATGTTCCAGAACTCT 115
   |||||.....:||||:||||:|||||
72 uValHisProGlnIleuTyrLeuThrAlaIaHisCysArgGlyLeuValP 89
   |||||.....:||||:||||:|||||
116 ACACATCGGGCTGGCGCTGCACAGTCTTGAGCCGACCAAGACCCAGG 165
   |||||.....:||||:||||:|||||
89 heArgValAlaGlyLeuGlnHisTyrIleuSerLeuSerProValTyrGluSerGly 105
   |||||.....:||||:||||:|||||
166 AGCAGATGATGGAGGAGCCAGCTCTCCGTACGAGCCAGACAGATACAG 215
   |||||.....:||||:||||:|||||
106 GlnGlnMetHegInGlnGlyValIysSerIleProHisProGlyTyrSerH 122
   |||||.....:||||:||||:|||||
216 ACCCTTGTGCTAGACATCATGTCATCAAGTTGAGAGAAATCCGCT 265
   |||||.....:||||:||||:|||||
122 sProGlnHisSerAsnAspLeuMetLeuIleLeuValAsnArgArgLeuA 139
   |||||.....:||||:||||:|||||
266 CCGAGTGTGACACATCCGAGCATCAGATGCTTGGCAGTCCCTACC 315
   |||||.....:||||:||||:|||||
139 rgrProThrIlyAspValAlaArgProIleAsnValSerSerHisCysProSer 155
   |||||.....:||||:||||:|||||
316 GCGGAGAACTTGTGCTGCTTGTGCTGGGAGTGTGCTGGGAGACGATGC 365
   |||||.....:||||:||||:|||||
156 AlaGlyThrIlyCysLeuValSerGlyTyrPheIlyThrIlySerProGln 172
   |||||.....:||||:||||:|||||
366 TGTTG.....ATTGCCATCCAGTCC 385
   |||||.....:||||:||||:|||||
172 nValHisPheProIlyValLeuGlnCysLeuAsnIleSerValLeuSerG 189
   |||||.....:||||:||||:|||||
386 AGACTGTGGAGGCTGGGAGTGTGAG 411
   |||||.....:||||:||||:|||||
189 InIys.....ArgCysGln 193
   |||||.....:||||:||||:|||||
```

seq_name: A_Geneseq_36:R67888

seq_documentation_block:
ID R67888 standard: Protein: 253 AA.
AC R67888, ..

DR 09-AUG-1995 (first entry)
DE Human stratum corneum chymotryptic recombinant enzyme (SCCE).
KW Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;
KW callosities; keratosis pilaris; ichthyoses; eczema.
OS Homo sapiens.
PN WO9500651-A.
PD 05-JAN-1995.
PF 20-JUN-1994; IB0166.
PR 18-JUN-1993; DK-000725.
PA (SYMB-) SYMBICOM AB.
PI Egelrud T, Hansson L;
DR N-PSDB; 081203.
DR Nucleotide sequences encoding stratum corneum chymotryptic enzyme
PT - and related vectors, transformed cells and polypeptides,
PT useful for treating skin disorders, e.g. acne or psoriasis, and
PS for identification of specific inhibitors.
PS Disclosure: Page 97; 137pp; English.
CC The enzyme encoded by this sequence is used in pharmaceutical, cosmetic
CC and skin care products, especially to treat and prevent acne,
CC xeroderma, or other hyperkeratotic conditions (e.g. callosities or
CC keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is
CC produced recombinantly following mammal, insect, plant, or
CC microorganism transformation with plasmid pS507.
SQ Sequence 253 AA;

alignment_scores: Quality: 274.50 Length: 143
 Ratio: 2.745 Gaps: 3
Percent Similarity: 69.930 Percent Identity: 38.462

alignment_block:
US-09-030-606-177 x R67888 ..

Align seg 1/1 to: R67888 from: 1 to: 253

```
7 TCGCAGCCCTGAGGAGCGGACATGTCATGAGAAACGAATGTTCTGTC 56
   |||||.....:||||:||||:|||||
40 SerHisProThrIleuValAlaLeuLeuSerGlyAsnGlnLeuHisCysG 56
   |||||.....:||||:||||:|||||
56 uGlyValLeuValAsnGlnArgTyrValLeuThrAlaIaHisCysGlyLeu 73
   |||||.....:||||:||||:|||||
107 AGAACCTTACACATCGGGCTGGCGCTGCACAGTCTTGAGCCGACCA 156
   |||||.....:||||:||||:|||||
73 etAsnGlnTyrThrValHisLeuGlySerAspThrLeu...GlyAspArg 88
   |||||.....:||||:||||:|||||
157 GACCCAGGAGCCAGATGGTGGAGGAGCCAGCTCTCCGTACGAGCCAGCA 206
   |||||.....:||||:||||:|||||
89 Arg.....AlaGlnArgIleuValAlaSerIlySerPheArgHisProGln 103
   |||||.....:||||:||||:|||||
207 GTACACAGACACCTTGTGCTAGACATCATGTCATCAAGTTGAGAGC 256
   |||||.....:||||:||||:|||||
103 uTyrSerThrGlnThrHisValAsnAspLeuMetLeuValIlyLeuAsn 120
   |||||.....:||||:||||:|||||
257 AATCCGTTGCGAGTGTGACACATCCGAGCATCAGATGCTTGGCAG 306
   |||||.....:||||:||||:|||||
120 erGlnAlaArgLeuSerSerMetValIlyValArgLeuProSerArg 136
   |||||.....:||||:||||:|||||
307 TGCCTTACCGCGGGAACCTTGTGCTGCTTGTGCTGGGAGTGTGCTGC 356
   |||||.....:||||:||||:|||||
137 CysGlnProProGlyThrThrCysThrValSerGlyTyrPheIlyThrThr 153
   |||||.....:||||:||||:|||||
357 GAACATGCTGTG.....ATTGCCATCCAGTCCAG 388
   |||||.....:||||:||||:|||||
153 rSerProAspValThrPheProSerAspLeuMetCysValAspValIlyL 170
   |||||.....:||||:||||:|||||
389 CTGTGGAGGCTGGAGTGTGAGAACTT 417
   |||||.....:||||:||||:|||||
170 euIleSerProGlnAspCysThrIlyVal 179
   |||||.....:||||:||||:|||||
```

seq_name: A_Geneseq_36:W05383

seq_documentation_block:

ID W05383 standard: Protein; 253 AA.
AC W05383;
DT 31-DEC-1996 (first entry)
DE Human amyloid precursor protein protease.
KW Amyloid precursor protein protease; Alzheimer's disease; diagnosis;
therapy
OS Homo sapiens.
PN W09631122-A1.
PD 10-OCT-1996.
PF 02-APR-1996; U04294.
PR 04-APR-1995; US-416257.
PA (ELIL) LILLY & CO ELI.
PI Dixon EP, Johnstone EM, Little SP;
DR MPI; 96-464694/46.
DR N-PSDB; T39783.
PT New isolated human amyloid precursor protein protease - used to
develop prods. for the treatment or diagnosis of associated
PT conditions, esp. Alzheimer's disease
PS Claim 1: Page 44-45; 55pp; English.
CC Human amyloid precursor protein protease (W05383) is involved in
CC the processing or clearance of amyloid precursor protein to form
CC beta-amyloid peptide. Its amino acid sequence was deduced from
CC a cDNA clone (T39783) obtd. from a human lung library. Recombinant
CC protease can be produced in transformed or transfected prokaryotic
CC (partic. E. coli) or eukaryotic (partic. AV-120 host cells. It is
CC used to develop products for the design and testing of cpds. useful
CC for treating or preventing conditions associated with beta-amyloid
CC peptide, esp. Alzheimer's disease.
SO Sequence 253 AA;

alignment_scores:

Quality: 274.50 Length: 143
Ratio: 2.745 Gaps: 3
Percent Similarity: 69.930 Percent Identity: 38.462

alignment_block:

US-09-030-606-177 x W05383 ..

Align seg 1/1 to: W05383 from: 1 to: 253

```

7  TCGAGCCCTGCGAGCGGCGACATGTCATGAAAACGATTTCTGCTC 56
|||||
40 SerHisProTrpGlnValAlaLeuLeuSerGlyAsnGlnLeuHisCysG1 56
57 GGGCGCTGCTGGATCCGACGTGGTGTGACCGCCGACACATGTTCC 106
|||||
56 YG1ValLeuValAsnGlnArgTrpValLeuThrAlaIleHisCysLysW 73
107 AGAAGCTCTACACCATCGCGCTGGGCTGACAGTCTTGAGCCGACCA 156
|||||
73 etAsnGlnTrpThrValHisLeuGlySerAspThrLeu...GlyAspArg 88
157 GAGCCAGGAGCCAGATGTGGAGCGCCCTCTCCGACGACCCGACGA 206
|||||
89 Arg.....AlaGlnArgIleLysAlaSerLysSerPheArgHisProG1 103
207 GTACACAGACCCCTTGCTGCTACGACCTCATGCTCATCAAGTTGAGAG 256
|||||
103 YTrpSerThrGlnThrHisValAsnAspLeuLeuValLysLeuAsnS 120
257 AATCCGTCGAGTCTGACACATCCGAGCATCCAGCATGCTTCGAG 306
|||||
120 eArgIleAlaArgLeuSerSerMetValLysLysValArgLeuProSerArg 136
307 TGCCTACCGCGGGGAACTTTGCTGCTTCTGCTGGGCTGCTGCTGCG 356
|||||
137 CysGlnProProGlyTrpThrCysThrValSerGlyTrpGlyThrThrTh 153
357 GAAGCATCTGTG.....ATTGCATTCAGTCGCCAGA 388

```

seq_name: A_Geneseq_36:W71005

seq_documentation_block:

ID W71005 standard: Protein; 262 AA.
AC W71005;
DT 19-OCT-1998 (first entry)
DE Human prostate-associated kallikrein designated HPAK.
KW Prostate-associated kallikrein; HPAK; human; pancreatic kallikrein;
KW prostate gland; development; cancer; antagonist; inhibitor;
KW suppression; excessive cell proliferation; treatment; prevention;
KW cancer; prostate hyperplasia; detection; diagnosis.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Modified_site 102 /note= "potential N-glycosylation site"
FT Modified_site 168 /note= "potential N-glycosylation site"
FT Modified_site 184 /note= "potential N-glycosylation site"
FT Modified_site 213 /note= "potential N-glycosylation site"
PN W09832865-A1.
PD 30-JUL-1998.
PF 26-JUN-1998; U01440.
PR 29-JAN-1997; US-790137.
PA (INCT-) INCTE PHARM INC.
PI Goli SR, Hillman JL;
DR N-PSDB; V42925.
PT New isolated prostate-associated kallikrein - is used to develop
PT products for the treatment of cancers, e.g. of the prostate, breast
PT or parotid gland, or other prostate disorders
PS Claim 1: Fig 1A-C; 59pp; English.
CC The present sequence represents a prostate-associated kallikrein protein
CC (HPAK). HPAK has chemical and structural homology with human pancreatic
CC kallikrein. HPAK contains conserved residues critical for serine
CC protease activity (His65, Asp113 and Ser206). Amino acid Asp200 is
CC likely to confer on HPAK chymotrypsinogen-like activity. The HPAK
CC protein sequence also contains 10 conserved Cys residues (31, 50, 66,
CC 145, 166, 177, 191, 202, 212 and 227). HPAK expression is associated with
CC the prostate gland and in the development of cancer. Antagonists or
CC inhibitors of HPAK may be used to suppress excessive cell proliferation.
CC They can be used to treat or prevent cancer, e.g. cancer of the prostate,
CC parotid gland and breast. They can also be used for other disorders of
CC the prostate, e.g. prostate hyperplasia. The products can also be used
CC for detection and diagnosis.
SO Sequence 262 AA;

alignment_scores:

Quality: 272.50 Length: 155
Ratio: 2.698 Gaps: 4
Percent Similarity: 65.161 Percent Identity: 38.710

alignment_block:

US-09-030-606-177 x W71005 ..

Align seg 1/1 to: W71005 from: 1 to: 262

```

4  CACGCGAGCCCTGCGAGCGGCGACATGTCATGAAAACGATTTCTG 53
|||||
34 HisSerGlnProTrpGlnAlaAlaLeuTrpHisPheSerThrPheGlnCys 50
54 CTCGCGGCTCTGCTGTCATCCGACGTGGTCTGTCAGCCGACACATGTT 103
|||||
50 sGlyGlyIleLeuValHisArgGlnTrpValLeuThrAlaIleHisCysI 67

```

```
104 TCCAGACTCTCTACACCATCGGGCTGGCCCTGCACAGCTTTGAGCCGAC 153
      :::::::::::::: :::: ||||| ||::::||| ||
67 leSeraspasnTYrGlnLeuTripleuGLyArGHisasnLeu...PheAsp 82
      ::::|
154 CAAGAGCCAGGAGCCAGATGTGGAGGCCAGCCCTCTCCGTACGGCACCC 203
      ::::| ::::| ::::| ::::| ::::| ::::|
83 AspGluasnThrAlaGlnPheValHisValSerGlnSerPheProHisPr 99
      ::::|
204 AGAGTCAACAGACCCCTTGCTCGCTAC..... 231
      | ::::| ||||| |||
99 OGlyPheasnMetSerLeuLeuGluasnHisThrArgGlnAlaAspGluA 116
      ::::|
232 .....GACCTCATGCTCATCAAGTTGGACGAATCCGTGCCGAG 270
      ||||| ||||| ::::| || ::::|
116 sPTyrSerHisAspLeuMetLeuLeuArgLeuThrGluProAlaAspThr 132
      ::::|
271 ...TCTGACACCATCCGAGAGCATCAGCATTGCTTGCAGTGCCCTACCGC 317
      ::::| ::::| ::::| ::::| ||| ::
133 IleThrAspAlaValAlaValValGluLeuProThrGlnGluProGluVa 149
      ::::|
318 GGGGAACTCTTGCTGCTTCTTGCTGGGGCTGCTGCGC..... 357
      ::::|
149 IGlySerThrCysLeuAlaSerGlyTrpGlySerIleGluProGluAsn 166
      ::::|
358 .....AAGCATGCTGTGATTGCCATCCAGTCCAGACTGTGGAGGC 399
      ||| ::::| ::::| ::::| ::::| ::::|
166 heSerPheProAspAspLeuGlnCysValAspLeuLysIleLeuProAsn 182
      ::::|
400 TGGGAGTGTGAGAAG 414
      |||||
183 AspGluCysGluLys 187
```

(TM)

Distribution rights by Oxford Molecular Ltd

in - protein database search, using Smith-Waterman algorithm

ned Sep 29 14:08:30 1999; MasPar time 11.40 Seconds

generated.

LDLQEASVADKLSFGKMAETRGTTWPHQGNHVRLLPR 37

PAM 150

179066 seqs, 54579741 residues

Listing first 45 summaries

sptremb19

13:sp_vertebrate 14:sp_virus

Mean 32.468; Variance 50.260; scale 0.646

ved by analysis of the total score distribution.

SUMMARIES

7.9	266 10 080337	ETHYLENE RESPONSIVE EL	8.14e-011
-----	---------------	------------------------	-----------

45	66	24.9	698	14	Q67889	DNA POLYMERASE (EC 2.7	1.58e+01
----	----	------	-----	----	--------	------------------------	----------

ALIGNMENTS

04 NOV 1950 (AMENDED: 00 / END) CANCELLATION OF DATE)
PERSEPHIN.

```

OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE: 98150950.
RA MILBRANDT J., DE SAUVAGE F.J., FAHRNER T.J., BALOH R.H., LEITNER M.L.,
RA TANSLEY M.G., LAMPE P.A., HEUCKEROH R.O., KOTZBAUER P.T.,
RA SIMBURGER K.S., GOLDEN J.P., DAVIES J.A., VESADA R., KATO A.C.,
RA HYNS M., SHERMAN D., NISHIMURA M., WANG L.-C., VANDLEN R., MOFFAT B.,
RA KLEIN R.D., POLSEN K., GRAY C., GARCES A., HENDERSON C.E.,
RA PHILLIPS H.S., JOHNSON E.M.,
RA "Persephin", a novel neurotrophic factor related to GDNF and
RT neurturin.
RT NEURON 20:245-253(1998).
DR EMBL: AF040961; G2935708;
SQ SEQUENCE 156 AA; 17063 MW; BA9BA08B CRC32;

Query Match
Best Local Similarity 73.2%; Score 194; DB 11; Length 156;
Matches 29; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

DB 24 LIDQEPADDELSSGKMAETRGTTWPHQNNHRLR 60
QY 1 LIDQEASVADKLSFGKMAETRGTTWPHQNNHRLR 37

RESULT 3
ID 031201; PRELIMINARY; PRT; 119 AA.
AC 031201;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE MHC H-2 S-REGION C4-SLP (SEX-LIMITED PROTEIN) (FRAGMENT).
OS MUS MUSCULUS (MOUSE)
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE: 85038859.
RA TOSI M., LEVI-STRAUSS M., DUPONCHEL C., MEO T.;
RT "Sequence heterogeneity of murine complementary DNA clones related to
RT the C4 and C4-SLP isoforms of the fourth complement component.";
RL PHILLOS. TRANS. R. SOC. LOND., B. BIOL. SCI. 306:389-394(1984).
DR EMBL: K02799; G199626;
KW PFAM: PF00207; A2M; 1.
MHC.
FT NON_TER 1 1
FT NON_TER 119 119
SQ SEQUENCE 119 AA; 12972 MW; 7302A10E CRC32;

Query Match
Best Local Similarity 30.6%; Score 81; DB 7; Length 119;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

DB 18 GKMAKRAASWTLHOGNFH 35
QY 15 GKMAETRGTTWPHQNNH 32

RESULT 4
ID 062238; PRELIMINARY; PRT; 594 AA.
AC 062238;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DE SEX-LIMITED PROTEIN SLP(W7) ALPHA-GAMA CHAIN (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RX SEQUENCE FROM N.A.
RX STRAIN-B10.D2(C4(H)SLP(A));

```

```

RX MEDLINE: 86176748.
RA HEMENWAY C., KALFF M., STAVENHAGEN J., WALTFRAL D., ROBINS D.;
RT "Sequence comparison of alleles of the fourth component of complement
RT (C4) and sex-limited protein (SLP).";
RL NUCLEIC ACIDS RES. 14:2539-2554(1986).
DR EMBL: X06454; G54106;
DR PFAM: PF00207; A2M; 1.
FT NON_TER 1 1
FT CHAIN 1 303
FT CHAIN 304 594
SQ SEQUENCE 594 AA; 66454 MW; C76E7A07 CRC32;

Query Match
Best Local Similarity 30.6%; Score 81; DB 11; Length 594;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

DB 131 GKMAKRAASWTLHOGNFH 148
QY 15 GKMAETRGTTWPHQNNH 32

RESULT 5
ID 082061; PRELIMINARY; PRT; 1464 AA.
AC 082061;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE PROTEIN INVOLVED IN STARCH METABOLISM PRECURSOR.
GN RL.
OS SOLANUM TUBEROSUM (POTATO).
OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLLIPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC ASTERIDAE; SOLANACEAE; SOLANALES; SOLANACEAE; SOLANUM.
RN [1]
RX SEQUENCE FROM N.A.
RX STRAIN-CV. DESIREE.
RA LOBERER R., KOSSMAN J.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: Y09533; E303790;
KW SIGNAL.
FT SIGNAL 1 48
FT SIGNAL 163236 163236
SQ SEQUENCE 1464 AA; 163236 MW; 564CF92A CRC32;

Query Match
Best Local Similarity 29.4%; Score 78; DB 10; Length 1464;
Matches 9; Conservative 15; Mismatches 12; Indels 2; Gaps 2;

DB 172 LEIRDTAI-EAIEFLIYDEAHDKWIKNGNFRVKLSR 208
QY 1 LIDQEASVADKLSFGKMAETRGTTWPHQNN-HVRLR 37

RESULT 6
ID 076253; PRELIMINARY; PRT; 344 AA.
AC 076253;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE TRANSPOSASE.
OS DROSOPHILA TEISSIERI (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MOSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RX SEQUENCE FROM N.A.
RX TRANSPOSON-MARINER-DTB24;
RX MEDLINE: 96270054.
RA BRUNET F., GODIN F., BAZIN C., DAVID J.R., CAPY P.;
RT "The mariner transposable element in natural populations of
RT Drosophila teissieri.";
RL J. MOL. EVOL. 42:669-675(1996).
RN [2]
RX SEQUENCE FROM N.A.

```


RC TRANSPOSON-MARINER-DT824;
 RA BRUNET F., GODIN F., CAPY P.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF035567; G3372339;
 SQ SEQUENCE 344 AA; 40553 MW; CC8E6AEA CRC32;

Query Match 29.1%; Score 77; DB 5; Length 344;
 Best Local Similarity 28.9%; Pred. No. 2,53e-01;
 Matches 11; Conservative 14; Mismatches 10; Indels 3; Gaps 3;

Db 96 LEVSOQAVSNRLREMGKT-QKVGRRVPHLENER-QMER 131
 1 IDLQASVADKLS-FGKMAETRGTTWPHOGNNHVRLPR 37

RESULT 7
 ID 061447; PRELIMINARY; PRT; 345 AA.
 AC 061447;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE TRANSPOSASE.
 OS DROSOPHILA SECELLIA (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
 OC DROSOPHILIDAE; DROSOPHILA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-228CSECH;
 RX MEDLINE: 92071978.
 RA CAPY P., MARUYAMA K., DAVID J.R., HARTL D.L.;
 RT "Insertion sites of the transposable element mariner are fixed in the genome of Drosophila sechellia."
 RL J. MOL. EVOL. 33:450-456(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-228CSECH;
 RX MEDLINE: 93106423.
 RA CAPY P., DAVID J.R., HARTL D.L.;
 RT "Evolution of the transposable element mariner in the Drosophila melanogaster species group."
 RL GENETICA 86:37-46(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-228CSECH;
 RX MEDLINE: 96270034.
 RA BRUNET F., GODIN F., BAZIN C., DAVID J.R., CAPY P.;
 RT "The mariner transposable element in natural populations of Drosophila teissieri."
 RL J. MOL. EVOL. 42:669-675(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-228CSECH;
 RX CAPY P.;
 RA SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF035569; G3136113;
 SQ SEQUENCE 345 AA; 40659 MW; 951E8393 CRC32;

Query Match 29.1%; Score 77; DB 5; Length 345;
 Best Local Similarity 28.9%; Pred. No. 2,53e-01;
 Matches 11; Conservative 14; Mismatches 10; Indels 3; Gaps 3;

Db 96 LEVSOQAVSNRLREMGKT-QKVGRRVPHLENER-QMER 131
 1 IDLQASVADKLS-FGKMAETRGTTWPHOGNNHVRLPR 37

RESULT 8
 ID 061451; PRELIMINARY; PRT; 345 AA.
 AC 061451;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE TRANSPOSASE.
 OS DROSOPHILA SIMULANS (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
 OC DROSOPHILIDAE; DROSOPHILA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-L14;
 RX MEDLINE: 93106423.
 RA CAPY P., KOGA A., DAVID J.R., HARTL D.L.;
 RT "Evolution of the transposable element mariner in the Drosophila melanogaster species group."
 RL GENETICA 86:37-46(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-L14;
 RX MEDLINE: 92201636.
 RA CAPY P., KOGA A., DAVID J.R., HARTL D.L.;
 RT "Sequence analysis of active mariner elements in natural populations of Drosophila simulans."
 RL GENETICS 130:499-506(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-L14;
 RX CAPY P.;
 RA SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF037055; G3136132;
 SQ SEQUENCE 345 AA; 40669 MW; F993F0BC CRC32;

Query Match 29.1%; Score 77; DB 5; Length 345;
 Best Local Similarity 28.9%; Pred. No. 2,53e-01;
 Matches 11; Conservative 14; Mismatches 10; Indels 3; Gaps 3;

Db 96 LEVSOQAVSNRLREMGKT-QKVGRRVPHLENER-QMER 131
 1 IDLQASVADKLS-FGKMAETRGTTWPHOGNNHVRLPR 37

RESULT 9
 ID 061453; PRELIMINARY; PRT; 345 AA.
 AC 061453;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE TRANSPOSASE.
 OS DROSOPHILA SIMULANS (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
 OC DROSOPHILIDAE; DROSOPHILA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-PRI;
 RX MEDLINE: 93106423.
 RA CAPY P., DAVID J.R., HARTL D.L.;
 RT "Evolution of the transposable element mariner in the Drosophila melanogaster species group."
 RL GENETICA 86:37-46(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-PRI;
 RX MEDLINE: 92201636.
 RA CAPY P., KOGA A., DAVID J.R., HARTL D.L.;
 RT "Sequence analysis of active mariner elements in natural populations of Drosophila simulans."
 RL GENETICS 130:499-506(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-PRI;
 RX CAPY P.;
 RA SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: AF037058; G3136138;
 SQ SEQUENCE 345 AA; 40768 MW; D50E4F24 CRC32;

Query Match 29.1%; Score 77; DB 5; Length 345;
 Best Local Similarity 28.9%; Pred. No. 2,53e-01;
 Matches 11; Conservative 14; Mismatches 10; Indels 3; Gaps 3;

Db 96 LEVSQAVSNRLREMGKI-QKGVWVPHLENER-QMER 131
 1 LDLOEASVADKLS-FGKMAETRGTTWPHGNNHVLP 37

RESULT 10 PRELIMINARY; PRT; 345 AA.

AC 061452;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE TRANSPOSASE.
 OS DROSOPHILA SIMULANS (FRUIT FLY).
 OC EURARCTOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
 OC DROSOPHILIDAE; DROSOPHILA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-MADB;
 RX MEDLINE; 93106423.
 RA CAPY P.; DAVID J.R.; HARTL D.L.;
 RT "Evolution of the transposable element mariner in the Drosophila melanogaster species group."
 RL GENETICA 86:37-46(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-MADB;
 RX MEDLINE; 92201636.
 RA CAPY P.; KOGA A.; DAVID J.R.; HARTL D.L.;
 RT "Sequence analysis of active mariner elements in natural populations of Drosophila simulans."
 RL GENETICS 130:499-506(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-MADB;
 RX MEDLINE; 96270054.
 RA BRUNET F.; GODIN F.; BAZIN C.; DAVID J.R.; CAPY P.;
 RT "The mariner transposable element in natural populations of Drosophila teissieri."
 RL J. MOL. EVOL. 42:669-675(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-MADB;
 RX CAPY P.;
 RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AF037056; G3136134; -;
 SO SEQUENCE 345 AA; 40789 MW; EFCFDDOF CRC32;

Query Match 29.1%; Score 77; DB 5; Length 345;
 Best Local Similarity 28.9%; Pred. No. 2,53e-01;

Matches 11; Conservative 14; Mismatches 10; Indels 3; Gaps 3;

Db 96 LEVSQAVSNRLREMGKI-QKGVWVPHLENER-QMER 131
 1 LDLOEASVADKLS-FGKMAETRGTTWPHGNNHVLP 37

RESULT 11 PRELIMINARY; PRT; 345 AA.

AC 061455;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE TRANSPOSASE.
 OS DROSOPHILA SIMULANS (FRUIT FLY).
 OC EURARCTOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
 OC DROSOPHILIDAE; DROSOPHILA.
 RN [1]-

RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-BORDA;
 RX MEDLINE; 93106423.
 RA CAPY P.; DAVID J.R.; HARTL D.L.;
 RT "Evolution of the transposable element mariner in the Drosophila melanogaster species group."
 RL GENETICA 86:37-46(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-BORDA;
 RX MEDLINE; 92201636.
 RA CAPY P.; KOGA A.; DAVID J.R.; HARTL D.L.;
 RT "Sequence analysis of active mariner elements in natural populations of Drosophila simulans."
 RL GENETICS 130:499-506(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-BORDA;
 RX CAPY P.;
 RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AF037060; G3136142; -;
 SO SEQUENCE 345 AA; 40774 MW; 065A5AAB CRC32;

Query Match 29.1%; Score 77; DB 5; Length 345;
 Best Local Similarity 28.9%; Pred. No. 2,53e-01;
 Matches 11; Conservative 14; Mismatches 10; Indels 3; Gaps 3;

Db 96 LEVSQAVSNRLREMGKI-QKGVWVPHLENER-QMER 131
 1 LDLOEASVADKLS-FGKMAETRGTTWPHGNNHVLP 37

RESULT 12 PRELIMINARY; PRT; 345 AA.

AC 062616;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE TRANSPOSASE.
 OS DROSOPHILA SIMULANS (FRUIT FLY).
 OC EURARCTOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
 OC DROSOPHILIDAE; DROSOPHILA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-SEY2, MARINER-MOS6A, MARINER-MOS6B;
 RX MEDLINE; 93106423.
 RA CAPY P.; DAVID J.R.; HARTL D.L.;
 RT "Evolution of the transposable element mariner in the Drosophila melanogaster species group."
 RL GENETICA 86:37-46(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-SEY2, MARINER-MOS6A, MARINER-MOS6B;
 RX MEDLINE; 92201636.
 RA CAPY P.; KOGA A.; DAVID J.R.; HARTL D.L.;
 RT "Sequence analysis of active mariner elements in natural populations of Drosophila simulans."
 RL GENETICS 130:499-506(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-SEY2, MARINER-MOS6A, MARINER-MOS6B;
 RX CAPY P.;
 RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AF037057; G3136136; -;
 DR EMBL; AF037052; G3136126; -;
 DR EMBL; AF037053; G3136128; -;
 SO SEQUENCE 345 AA; 40798 MW; DEL5C886 CRC32;

Query Match 29.1%; Score 77; DB 5; Length 345;
 Best Local Similarity 28.9%; Pred. No. 2,53e-01;
 Matches 11; Conservative 14; Mismatches 10; Indels 3; Gaps 3;

Db 96 LEVSOQAVSNRLREMGRI-QKVGWVPHELNER-OMER 131
 1 LDLOEASVADKLS-FGKMAETRGWTWPHOGNNHYRLPR 37

OY 13 PRELIMINARY; PRT; 345 AA.

RESULT 13
 ID 061454 PRELIMINARY; PRT; 345 AA.
 AC 061454;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE TRANSPOSASE.
 OS DROSOPHILA SIMULANS (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
 OC DROSOPHILIDAE; DROSOPHILA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-PA2SECH;
 RX CAPY P., MARUYAMA K., DAVID J.R., HARTL D.L.;
 RT "insertion sites of the transposable element mariner are fixed in the
 genome of Drosophila sechellia.";
 RL J. MOL. EVOL. 33:450-456(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-PA2SECH;
 RX MEDLINE; 93106423.
 RA CAPY P., DAVID J.R., HARTL D.L.;
 RT "Evolution of the transposable element mariner in the Drosophila
 melanogaster species group.";
 RL GENETICA 86:37-46(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-PA2;
 RX MEDLINE; 92201636.
 RA CAPY P., KOGA A., DAVID J.R., HARTL D.L.;
 RT "Sequence analysis of active mariner elements in natural populations
 of Drosophila simulans.";
 RL GENETICS 130:499-506(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-PA2;
 RX MEDLINE; 92201636.
 RA CAPY P., KOGA A., DAVID J.R., HARTL D.L.;
 RT "Sequence analysis of active mariner elements in natural populations
 of Drosophila simulans.";
 RL GENETICS 130:499-506(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-PA2;
 RX MEDLINE; 92201636.
 RA CAPY P., KOGA A., DAVID J.R., HARTL D.L.;
 RT "Sequence analysis of active mariner elements in natural populations
 of Drosophila simulans.";
 RL GENETICS 130:499-506(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-PA2;
 RX MEDLINE; 92201636.
 RA CAPY P., KOGA A., DAVID J.R., HARTL D.L.;
 RT "Sequence analysis of active mariner elements in natural populations
 of Drosophila simulans.";
 RL GENETICS 130:499-506(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-PA2SECH;
 RX CAPY P., MARUYAMA K., DAVID J.R., HARTL D.L.;
 RT "insertion sites of the transposable element mariner are fixed in the
 genome of Drosophila sechellia.";
 RL J. MOL. EVOL. 33:450-456(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-PA2SECH;
 RX MEDLINE; 93106423.
 RA CAPY P., DAVID J.R., HARTL D.L.;

Query Match 29.1%; Score 77; DB 5; Length 345;
 Best Local Similarity 28.9%; Pred. No. 2.53e-01;
 Matches 11; Conservative 14; Mismatches 10; Indels 3; Gaps 3;

Db 96 LEVSOQAVSNRLREMGRI-QKVGWVPHELNER-OMER 131
 1 LDLOEASVADKLS-FGKMAETRGWTWPHOGNNHYRLPR 37

RESULT 14
 ID 061446 PRELIMINARY; PRT; 345 AA.

AC 061446;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE TRANSPOSASE.
 OS DROSOPHILA SECHELLIA (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
 OC DROSOPHILIDAE; DROSOPHILA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-PA2SECH;
 RX MEDLINE; 92071978.
 RA CAPY P., MARUYAMA K., DAVID J.R., HARTL D.L.;
 RT "insertion sites of the transposable element mariner are fixed in the
 genome of Drosophila sechellia.";
 RL J. MOL. EVOL. 33:450-456(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-PA2SECH;
 RX MEDLINE; 93106423.
 RA CAPY P., DAVID J.R., HARTL D.L.;

RT "Evolution of the transposable element mariner in the Drosophila
 melanogaster species group.";
 RL GENETICA 86:37-46(1992).

RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-PA2SECH;
 RX MEDLINE; 96270054.
 RA BRUNET F., GODIN F., BAZIN C., DAVID J.R., CAPY P.;
 RT "The mariner transposable element in natural populations of
 Drosophila teissieri.";
 RL J. MOL. EVOL. 42:669-675(1996).

RN [4]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-PA2SECH;
 RX CAPY P.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AF035568; G3136111; -;
 SO SEQUENCE 345 AA; 40768 MW; 6C1D83C8 CRC32;

Query Match 29.1%; Score 77; DB 5; Length 345;
 Best Local Similarity 28.9%; Pred. No. 2.53e-01;
 Matches 11; Conservative 14; Mismatches 10; Indels 3; Gaps 3;

Db 96 LEVSOQAVSNRLREMGRI-QKVGWVPHELNER-OMER 131
 1 LDLOEASVADKLS-FGKMAETRGWTWPHOGNNHYRLPR 37

RESULT 15
 ID 061450 PRELIMINARY; PRT; 345 AA.

AC 061450;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE TRANSPOSASE.
 OS DROSOPHILA SIMULANS (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
 OC DROSOPHILIDAE; DROSOPHILA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-L8;
 RX MEDLINE; 93106423.
 RA CAPY P., DAVID J.R., HARTL D.L.;
 RT "Evolution of the transposable element mariner in the Drosophila
 melanogaster species group.";
 RL GENETICA 86:37-46(1992).

RN [2]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-L8;
 RX MEDLINE; 92201636.
 RA CAPY P., KOGA A., DAVID J.R., HARTL D.L.;
 RT "Sequence analysis of active mariner elements in natural populations
 of Drosophila simulans.";
 RL GENETICS 130:499-506(1992).

RN [3]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-L8;
 RX BRUNET F., GODIN F., BAZIN C., DAVID J.R., CAPY P.;
 RT "The mariner transposable element in natural populations of
 Drosophila teissieri.";
 RL J. MOL. EVOL. 42:669-675(1996).

RN [4]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-MARINER-L8;
 RX CAPY P.;
 RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AF037054; G3136130; -;
 SO SEQUENCE 345 AA; 40788 MW; 435FE906 CRC32;

Query Match 29.1%; Score 77; DB 5; Length 345;
 Best Local Similarity 28.9%; Pred. No. 2.53e-01;

Matches 11; Conservative 14; Mismatches 10; Indels 3; Gaps 3;

DB 96 LEVSOQAVSNRLREMGKI-QKYGWVPHELNER-QMER 131

QY 1 LDLOEASVADKLS-FKKMAETRGITWPHOGNNHVLPR 37

Search completed: Wed Sep 29 14:08:46 1999
Job time : 16 secs.

JOURNAL Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th st SE,
Bothell, WA 98021, USA
Location/Qualifiers
FEATURES
Source 1..1347

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13"
1..765
/note="androgen regulated"

/codon_start=1
/product="serine protease precursor"
/protein_id="AAB21580.1"
/db_xref="PID:94512030"
/db_xref="GI:4512030"

/translation="MTAGNPMGWFLLGILGSLAFINGHILNGEDCNPHSOPM
OALVVENELFCSCGVLVHPQWVLSAHCFTONSTTIGLHLSLEADPEPSQWVSLAS
VRHPEYNRLPLANDLMLRLKLSVLSSTIRISISQCPVTAQNSCLVSGWGLANGR
MPTVLQCVNIVSAVESEVSKLYDPLVHPSMFCAGGHDOKDSNDSGGLICNGYLQ
GLVSGKAPCGVGVGVYTNLCKFTMEIKTVQAS"

POLYA.site 1327..1332
BASE COUNT 269 a 489 c 334 g 255 t
ORIGIN

Query Match 32.2%; Score 360.8; DB 42; Length 1347;
Best Local Similarity 96.8%; Pred. No. 1.2e-91;

Matches 368; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 2 CGCAGCTGCAAGCCCTGGCAGGCGGCACTGTCATGAAAGCAATTGTTGCTGGGCG 61

DB 116 CGCAGCTGCAAGCCCTGGCAGGCGGCACTGTCATGAAAGCAATTGTTGCTGGGCG 175

QY 62 TCCGTGTCATCCGCACTGGGTGCTGTCACAGCCGACACTGTTCCAGAACTCCACACCA 121

DB 176 TCCGTGTCATCCGCACTGGGTGCTGTCACAGCCGACACTGTTCCAGAACTCCACACCA 235

QY 122 TCGGGCTGGGCTGTCACAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGGTGAGG 181

DB 236 TCGGGCTGGGCTGTCACAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGGTGAGG 295

QY 182 CCAGCTCTCCGTACGGCACCACCAAGATACACAGACCTTGCTGCTAACGACCTCATGC 241

DB 296 CCAGCTCTCCGTACGGCACCACCAAGATACACAGACCTTGCTGCTAACGACCTCATGC 355

QY 242 TCATCAAGTTGGAGCAATCCGTGCGAGTCTGACCAATCCGAGATCAGATTCCTT 301

DB 356 TCATCAAGTTGGAGCAATCCGTGCGAGTCTGACCAATCCGAGATCAGATTCCTT 415

QY 302 CGCAGTGGCCCTACCGGCGGGAACCTTGCTGCTGCTGGGTCTGCTGGCGAAGC 361

DB 416 CGCAGTGGCCCTACCGGCGGGAACCTTGCTGCTGCTGGGTCTGCTGGCGAAGC 475

QY 362 ATGCTGTGATGCCATCCAG 381

DB 476 GCAGAAATGCTACCGTCTG 495

RESULT 2

SSU76256 1140 bp mRNA MAM 03-JAN-1998

LOCUS Sus scrofa enamel matrix serine proteinase 1 precursor, mRNA,
DEFINITION complete cds.

ACCESSION U76256

VERSION 92737920

KEYWORDS GI:2737920

SOURCE Sus scrofa

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE Eutheria; Cetartiodactyla; Suidae; Sus.
AUTHORS 1 (bases 1 to 1140)
Stimmer,J.P., Fukae,M., Tanabe,T., Yamakoshi,Y., Uchida,T., Xue,J.,

Margolis,R.C., Shimizu,M., Hu,C.-C. and Bartlett,J.D.
Purification, Characterization and Cloning of Enamel Matrix Serine
Proteinase 1

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1140)

AUTHORS Stimmer,J.P., Fukae,M., Tanabe,T., Yamakoshi,Y., Uchida,T., Xue,J.,
Margolis,R.C., Shimizu,M., Hu,C.-C. and Bartlett,J.D.

TITLE Direct Submission

JOURNAL Submitted (24-OCT-1996) Pediatric Dentistry, University of Texas
Health Science Center at San Antonio, 7703 Floyd Curl Drive, San
Antonio, TX 78284-7888, USA

Location/Qualifiers

1..1140

/organism="Sus scrofa"

/db_xref="taxon:9823"

sig_peptide

69..140

/note="the signal peptide is cleaved after Ala24 and the
prepeptide is secreted into the developing enamel matrix"

69..833

/note="EMSP1"

/codon_start=1

/product="enamel matrix serine proteinase 1 precursor"

/protein_id="AAB94638.1"

/db_xref="PID:92737921"

/db_xref="GI:2737921"

/translation="MTAGNPMGWFLLGILGSLAFINGHILNGEDCNPHSOPM
OALVVENELFCSCGVLVHPQWVLSAHCFTONSTTIGLHLSLEADPEPSQWVSLAS
VRHPEYNRLPLANDLMLRLKLSVLSSTIRISISQCPVTAQNSCLVSGWGLANGR
MPTVLQCVNIVSAVESEVSKLYDPLVHPSMFCAGGHDOKDSNDSGGLICNGYLQ
GLVSGKAPCGVGVGVYTNLCKFTMEIKTVQAS"

159..830

/note="the prepeptide is cleaved following His30
activating the serine proteinase; the active protein has
an apparent molecular weight of 34 kDa and a derived
molecular weight of 24 kDa"

/product="enamel matrix serine proteinase 1"

order(177..179,549..551)

/note="disulfide bond"

order(234..236,282..284)

/note="disulfide bond"

join(279..281,414..416,687..689)

/note="encodes catalytic triad"

405..407

/note="encodes potential glycosylation location; yields a
blank cycle during protein sequencing"

order(408..410,789..791)

/note="disulfide bond"

order(510..512,705..707)

/note="disulfide bond"

order(675..677,750..752)

/note="disulfide bond"

FEATURES

source

sig_peptide

CDS

mat_peptide

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

Db 364 CCAGCCTCTCCATTCAGACCCAGAAATACAGAGCCCTCTATGCGCAACGACCTCATGTC 423
QY 242 TCATCAAGTTGGAGACATCCGTTGTCGAGTCTGACACCATCGAGCATGACATTCGTT 301
Db 424 TCATCAAGTTGAAGAAATCGGTGCTGTCTGACACCGTCCGAAACATCAACGCTGCT 483
QY 302 CGCAGTGGCCCTACCGCGGGGAACTTTCCTGTTCTGCTGGGCTGCTGGCGCA 358
Db 484 CCCAGTGGCCGACCCCTGGGATTCCTCTGCTGCTGGGCTGGGCTGGCGCA 540

RESULT 3
AF113141 5900 bp DNA PRI 25-MAR-1999
LOCUS Homo sapiens serine protease prostate gene, complete cds.
DEFINITION AF113141
ACCESSION AF113141
NID 94512031
VERSION AF113141.1 GI:4512031
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 5900)
Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gellinas,R., Hood,L. and Wang,K.
Molecular cloning and characterization of prostate, an androgen-regulated serine protease with prostate-restricted expression
Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)
MEDLINE 99179024
REFERENCE 2 (bases 1 to 5900)
Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gellinas,R., Hood,L. and Wang,K.
Direct Submission
Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th St SE, Bothell, WA 98021, USA
JOURNAL location/Qualifiers
TITLE 1. 5900
AUTHORS /organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13"
JOURNAL join(<1411..1471,2735..2897,3319..3569,3653..3789,5062..5817)
REFERENCE /product="serine protease prostate"
1411..1471
JOURNAL /number=1
exon join(1411..1471,2735..2897,3319..3569,3653..3789,5062..5214)
CDS /note="androgen regulated"
/codon_start=1
/product="serine protease prostate"
/protein_id="AAD21581.1"
/db_xref="PID:94512032"
/db_xref="GI:4512032"
translation="MARAGPMKMFLLGLVAGSLVSSCSQIINGEDCSPHSQPM QALVMENELFCGSLVHPQWLSAAHCFONSTTIGLHSLHLEDDPGSQMVEASIS VRRPEYRPLANDLMLIKLDESVESEDTRISISIAQCPTAGNSCLVSGMGLANR MPVLQCVNVSVAEEVCSKLDPLVHPSMFCAGGGGCHDQDSCNGDSGGGLICNGYLQ GLVSGKAPCGGGOVPGVYTNLCFETMIKTYOAS"
2735..2897
/number=2
exon 3319..3569
/number=3
exon 3653..3789
/number=4
exon 5062..5817
/number=5
BASE COUNT 1417 a 1643 c 1569 g 1271 t
ORIGIN

Query Match 22.8%; Score 255.2; DB 42; Length 5900;
Best Local Similarity 98.8%; Pred. No. 1.1e-61;
Matches 257; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 105 CCAGACTCTCTACACCATCGGGCTGGGCTGACACAGTCTTGAGGCGGACCAAGACCAGG 164
Db 3313 CCACAGCTCTCTACACCATCGGGCTGGGCTGACACAGTCTTGAGGCGGACCAAGACCAGG 3372
QY 165 GAGCAGATGGTGAAGCCAGCCCTCTCCGACGACACAGATACAGACACCTTGCT 224
Db 3373 GAGCAGATGGTGAAGCCAGCCCTCTCCGACGACACAGATACAGACACCTTGCT 3432
QY 225 CGCTAAGACCTCTGCTCATCAAGTTGAGAGAAATCCGTGCGAGTCTGACACATCCG 284
Db 3433 CGCTAAGACCTCTGCTCATCAAGTTGAGAGAAATCCGTGCGAGTCTGACACATCCG 3492
QY 285 GAGCATCAGCATTTGCTTGCAGTGCCTACCGCGGGGAACTTTCCTGCTTCTGCTG 344
Db 3493 GAGCATCAGCATTTGCTTGCAGTGCCTACCGCGGGGAACTTTCCTGCTTCTGCTG 3552
QY 345 GGGTCTGCTGGCGAACCATG 364
Db 3553 GGGTCTGCTGGCGAACCATG 3572

RESULT 4
AF135023 4740 bp DNA PRI 20-APR-1999
LOCUS Homo sapiens kallikrein-like protein 1 KRL-L1 gene, partial cds.
DEFINITION AF135023
ACCESSION AF135023
NID 94589272
VERSION AF135023.1 GI:4589272
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 4740)
Yousef,G.M., Luo,L.Y. and Diamandis,E.P.
Identification of novel human kallikrein-1-like genes on chromosome 2 (bases 1 to 4740)
Biochem. Biophys. Res. Commun. (1999) In press
JOURNAL Yousef,G.M., Luo,L.Y. and Diamandis,E.P.
REFERENCE 2 (bases 1 to 4740)
JOURNAL Direct Submission
Submitted (13-MAR-1999) Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada
location/Qualifiers
TITLE 1. 4740
AUTHORS /organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.3-q13.4"
JOURNAL join(<2263..2425,2847..3097,3181..3317,4588..>4740)
REFERENCE /product="kallikrein-1-like protein 1 KRL-L1"
join(<2263..2425,2847..3097,3181..3317,4588..>4740)
JOURNAL /codon_start=3
exon /product="kallikrein-1-like protein 1 KRL-L1"
/protein_id="AAD26424.1"
/db_xref="PID:94589273"
/db_xref="GI:4589273"
translation="SLVSGSCSQIINGEDCSPHSQPMQALVMENELFCGSLVHPQW VLSAAHCFONSTTIGLHSLHLEDDPGSQMVEASISVRRPEYRPLANDLMLIKLDESVESEDTRISISIAQCPTAGNSCLVSGMGLANRMPVLQCVNVSVAEEVCSKLDPLVHPSMFCAGGGGCHDQDSCNGDSGGGLICNGYLQGLVSGKAPCGGGOVPGVYTN LCKFETMIKTYOAS"
CDS
BASE COUNT 1106 a 1275 c 1308 g 1051 t
ORIGIN

Query Match 22.8%; Score 255.2; DB 42; Length 4740;
Best Local Similarity 98.8%; Pred. No. 1e-61;

	Matches	257,	Conservative	0;	Mismatches	3;	Indels	0;	Gaps	0
QY	105	CCAGAAACCTCTACCACTATGCGGGCTGGGCGCTGCACACTCTTGAAGCCGACCAAGACCCAG	164							
Db	2841	CCACAGCTCTCTACACCATCGGGCTGGGCTGCACACTCTTGAAGCCGACCAAGACCCAG	2900							
QY	165	GAGCCAGATGTGTGAGAGCCAGCCTCTCCCTAGCGCACCCAGATACACAGACCTTGGT	224							
Db	2901	GAGCCAGATGTGTGAGAGCCAGCCTCTCCCTAGCGCACCCAGATACACAGACCTTGGT	2960							
QY	225	CGCTAACGACCTCATAGCTCATATCAAGTTGAGCAAAATCCGTGCGGATCTGACACCAATCCG	284							
Db	2961	CGCTAACGACCTCATAGCTCATATCAAGTTGAGCAAAATCCGTGCGGATCTGACACCAATCCG	3020							
QY	285	GAGCATACAGCATTTGCTTCGACAGTGCCCTACCGCGGGGAACTCTTGCTCGTTTCTGGCTG	344							
Db	3021	GAGCATACAGCATTTGCTTCGACAGTGCCCTACCGCGGGGAACTCTTGCTCGTTTCTGGCTG	3080							
QY	345	GGGTCTGCTGGGGGAAGCATG	364							
Db	3081	GGGTCTGCTGGGGGAAGCATG	3100							

RESULT		5							
AF148532									
LOCUS	AF148532		4385 bp	DNA		PRI	09-JUN-1996		
DEFINITION	Homo sapiens kallikrein 4 (KLK4) gene, complete cds.								
ACCESSION	AF148532								
NID	G5020095								
VERSION	AF148532.1		GI:5020095						
KEYWORDS	.								
SOURCE	human.								
ORGANISM	Homo sapiens								

REFERENCE AUTHORS TITLE	JOURNAL REFERENCE AUTHORS TITLE
1 (bases 1 to 4385) Stephenson, S.A., Verity, K., Ashworth, L. and Clements, J.A. Localization of a new prostate specific antigen-related serine protease gene, KLK4, is evidence for an expanded human kallikrein (K1K) gene family cluster on chromosome 19q13.3-13.4 unpublished	2 (bases 1 to 4385) Stephenson, S.A., Ashworth, L. and Clements, J.A. Direct Submission Submitted (03-MAY-1999) School of Life Science, Queensland University of Technology, George Street, Brisbane, Queensland 4001 Australia

FEATURES	Location/Qualifiers
source	1..4385
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="19"
	/map="19q13.3-q13.4"
mrna	join(551..611,1875..2037,2459..2709,2793..2929,4200..>4352)
	/gene="KLU4"
	/product="kalikrein 4"
gene	<551..>4352
	/gene="KLU4"
CDS	join(551..611,1875..2037,2459..2709,2793..2929,4200..>4352)

```

/product="kallikrein 4"
/protein_id="AAD3801.1"
/db_xref="PID:95020096"
/db_xref="GI:5020096"
/translation="MATAGNPMWFLGYLLIGVAGSLVSGSCQIINGEDCSPHSPQPM
OAAALYMENELFCGYLIPWQVLSAAGFQNSYITGLDLHLEADQEGSQMVAASLS
VRHREYNRPCLLADMLTKIDSESESTISISALCPAGNSCLVSGGLANG
MPTLYACVNSVSVSEVSCSLYDPLIYHSMFCAGGQDQKSCNDSDGPIILNGIYL
GLVSGKAPCGQGYRPGYIKLDTFTIEKTVQAS"

```

ORIGIN

Query Match	22.8%;	Score 255.2;	DB 42;	Length 4385;
Best Local Similarity	98.8%;	Pred. No. 1e-61;		
Matches 257; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

OY	105	CCAAACATCCCTACACCAATCGGGCTGGGGCCGACACATCTTGAAGGCGGACCAAGACACAG	164
Db	2453	CCACAGCTCCCAACCAATCGGGGCTGGGGCCCTGCACACATCTTGAAGCGGACCAAGAGACAG	2512
OY	165	GAGCCAGATGTGTGAGGCCACGCTCTTCGTAGGGCACCCAGAGTACACAGACCCCTGCT	224
Db	2513	GAGCCAGATGTGTGAGGCCACGCTCTTCGTAGGCCACCCAGAGTACACAGACCCCTGCT	2572
OY	225	CGCTAACGACCTCATGTGTATCAAGTTGGAGCAATCCGTGCCAGTCTGACACACATCCG	284
Db	2573	CGCTAACGACCTCATGTCTCATCAAGTTGGAGCAATCCGTGTCGAGTCTGACACACATCCG	2632
OY	285	GAGCATCAGCATTTGCTTCGCGAGTGCCTACCGCGGGGAACTCTTGCTGTTTCGGCTG	344
Db	2633	GAGCATCAGCATTTGCTTCGCGAGTGCCTACCGCGGGGAACTCTTGCTGTTTCGGCTG	2692
OY	345	GGGTCCTGCTGGCGGAACGATG	364
Db	2693	GGGTCCTGCTGGCGGAACGATG	2712

RESULT	6				
AF019979					
LOCUS	AF019979	1237 bp	mRNA	ROD	01-JAN-1999
DEFINITION	Mus musculus enamel matrix proteinase 1 precursor, mRNA.				
ACCESSION	AF019979	complete cds.			
NTID	94090846				
VERSION	AF019979.1	GI:4090846			
KEYWORDS					
SOURCE	house mouse.				
ORGANISM	Mus musculus				

REFERENCE	1 (bases 1 to 1237)
AUTHORS	Simmer, J.
TITLE	Enamel Matrix Serine Proteinase 1 (EMSP1)
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1237)
AUTHORS	Simmer, J.
TITLE	Direct Submission
JOURNAL	Submitted (18-AUG-1997) Pediatric Dentistry, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78284-7888, USA
FEATURES	Location/Qualifiers

```

source
1. .1237
/organism="Mus musculus"
/strain="Swiss-Webster"
/db_xref="taxon:10090"
/cell_type="ameloblast"
52. .126
sig_peptide
CDS
52. .819
/function="degradation of the enamel matrix during enamel
maturation"
/note="EMSP1"
/codon_start=1
/product="enamel matrix serine proteinase 1 precursor"
/protein_id="AAC38894.1"
/db_xref="PID:94090847"
/db_xref="GI:4090847"
/translition="MAMVARTPTMGWFLGLILEVTGASASSVSRIIOGDGCSHSP
WQAAAFSEDFGFCGLVHPQWVLSAAHCLQESTYIVGLNLKSGSPGRMEAH
SHQHNEDSPRANMILIKINSEYSEINTTRSPVATOCPTPGDGLVSGWGLKNG
KLPSLQCVNLSVASSEETCRLLIDPVYHLLNSFCAGGGGDDOCDSCNGSDGPIYCNRSI
QGLVMSGQKCGQPEIPFSVYNLCKFTWMIOTITQN"
127. .144
msc_feature

```


repeat_region	/note="LI repeat: matches 5015. .4754 of consensus"
repeat_region	/note="LIM44 repeat: matches 1047. .801 of consensus"
repeat_region	/note="LI repeat: matches 2529. .2371 of consensus"
repeat_region	/note="LI repeat: matches 2203. .1103 of consensus"
repeat_region	/note="MER25 repeat: matches 2136. .1177 of consensus"
repeat_region	/note="MER25 repeat: matches 1031. .582 of consensus"
repeat_region	/note="MER25 repeat: matches 205. .13 of consensus"
repeat_region	/note="MST1 repeat: matches 384. .426 of consensus"
repeat_region	/note="MST-INTERNAL repeat: matches 1. .394 of consensus"
repeat_region	/note="LIMB1 repeat: matches 914. .757 of consensus"
misc_feature	23755..23876
repeat_region	/note="match: SRS L41276"
repeat_region	23842..24139
repeat_region	/note="ALUSC repeat: matches 1. .299 of consensus"
repeat_region	/note="ALUdb repeat: matches 293. .7 of consensus"
repeat_region	/note="THELB repeat: matches 364. .1 of consensus"
repeat_region	/note="THELB-INTERNAL repeat: matches 1580. .471 of consensus"
repeat_region	/note="THELC repeat: matches 44. .3 of consensus"
repeat_region	/note="ALUSg repeat: matches 299. .1 of consensus"
repeat_region	/note="MERSB repeat: matches 174. .8 of consensus"
repeat_region	28721..28894
repeat_region	/note="MIR repeat: matches 82. .258 of consensus"
repeat_region	29430..30796
repeat_region	/note="LI repeat: matches 4045. .5390 of consensus"
repeat_region	30659..31555
repeat_region	/note="LIPAL6 repeat: matches 1. .904 of consensus"
repeat_region	31655..31648
repeat_region	/note="12 copies 2 mer tt 100% conserved"
repeat_region	/note="ALVya5 repeat: matches 301. .1 of consensus"
repeat_region	/note="LIMD1 repeat: matches 971. .806 of consensus"
repeat_region	/note="MER42C repeat: matches 1338. .1485 of consensus"
repeat_region	/note="MER42C repeat: matches 1301. .173 of consensus"
repeat_region	/note="MER42C repeat: matches 13214. .33525
repeat_region	/note="LIMC3 repeat: matches 1362. .1054 of consensus"
repeat_region	/note="LIM48 repeat: matches 617. .3 of consensus"
repeat_region	/note="LI repeat: matches 33931. .36173
repeat_region	/note="LI repeat: matches 5390. .5158 of consensus"
repeat_region	36176..36573
repeat_region	/note="LIM49 repeat: matches 620. .1016 of consensus"
repeat_region	/note="ALU repeat: matches 292. .1 of consensus"
repeat_region	/note="LI repeat: matches 4630. .3154 of consensus"
repeat_region	/note="LI repeat: matches 3086. .2235 of consensus"
repeat_region	/note="LIM43 repeat: matches 148. .61 of consensus"
repeat_region	/note="LI repeat: matches 5287. .4843 of consensus"
repeat_region	44097..44291
repeat_region	/note="MIR repeat: matches 1. .204 of consensus"
repeat_region	complement(44511..44801)

[illegible]

SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 (bases 1 to 61450)
AUTHORS
Phillips,S.
TITLE
Direct Submission
JOURNAL
Submitted (07-DEC-1998) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Dec 2, 1998 this sequence version replaced gi:3676174.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
IMPORTANT: This sequence is not the entire insert of clone 15D7. It
may be shorter because we only sequence overlapping sections once,
or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone 3418 (AL021918) is at 6151. In this
sequence. The true right end of clone 97D16 (AL009179) is at 100 in
this sequence. This sequence has been finished according to
sequence map criteria as follows. An attempt is made to resolve all
sequencing problems, such as compressions and repeats, but not
necessarily within known annotated human repeat sequence elements
(e.g. Alu). Where the sequence is ambiguous, there is an annotation
using the 'unsure' feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed in collaboration with the Sanger
Centre chromosome 6 mapping group and Armin Volz & Andreas Ziegler,
David Ruddy, Jeffrey Green. Further information can be found at
<http://www.sanger.ac.uk/BCP/chr6/15D7>
15D7 is from the library RPC11 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see <http://daccpac.med.buffalo.edu/VCTOR:pcypac2>.
FEATURES
source
1..61450
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="15D7"
/map="p22.1-22.3"
/clone.lib="RPC11"
381..821
/note="match: GSS AQ213610"
643..950
repeat_region
/note="AluSp repeat: matches 1..309 of consensus"
1312..1385
/note="tRNA-Arg-CGC repeat: matches 1..74 of consensus"
1369..1486
/note="match: STS 278001"
complement(1494..1797)
/note="match: STS G02930"
1960..2001
/note="21 copies 2 mer ag 76% conserved"
1968..2003
repeat_region
/note="9 copies 4 mer agac 92% conserved"
2803..3107
repeat_region
/note="AluSq repeat: matches 1..308 of consensus"
3292..3366
/note="tRNA-Ile-ATT repeat: matches 1..75 of consensus"
3673..3730
/note="MER66A repeat: matches 284..340 of consensus"
5345..5643
repeat_region
/note="AluSx repeat: matches 10..310 of consensus"
5751..5899
repeat_region
/note="FLAM_A repeat: matches 1..125 of consensus"
5903..5954
repeat_region
/note="26 copies 2 mer tg 79% conserved"
5927..5954
repeat_region
/note="7 copies 4 mer tgtg 100% conserved"

misc_feature
6100..6268
/note="match: GSS B55592"
repeat_region
6208..6511
/note="AluSp repeat: matches 1..296 of consensus"
repeat_region
6564..6804
/note="AluY repeat: matches 1..301 of consensus"
7108..7183
/note="tRNA-Phe-TTC repeat: matches 1..73 of consensus"
7205..7250
/note="23 copies 2 mer tt 96% conserved"
repeat_region
7466..7791
/note="AluSp repeat: matches 1..313 of consensus"
8151..8461
/note="AluY repeat: matches 3..304 of consensus"
8462..8579
repeat_region
/note="MER92C repeat: matches 200..311 of consensus"
8580..8664
repeat_region
/note="MER68A repeat: matches 1..84 of consensus"
8785..8920
repeat_region
/note="MER92C repeat: matches 414..552 of consensus"
9471..9769
repeat_region
/note="AluSx repeat: matches 1..300 of consensus"
9773..9804
repeat_region
/note="16 copies 2 mer tt 91% conserved"
9882..10170
repeat_region
/note="AluJo repeat: matches 1..297 of consensus"
10186..10368
repeat_region
/note="L1MC2 repeat: matches 6158..6325 of consensus"
10370..10661
repeat_region
/note="AluSg repeat: matches 1..307 of consensus"
10676..10978
repeat_region
/note="AluSg repeat: matches 1..308 of consensus"
10991..11687
repeat_region
/note="L1MC2 repeat: matches 5173..5870 of consensus"
11682..12235
repeat_region
/note="L1MC4 repeat: matches 663..1295 of consensus"
12236..12548
repeat_region
/note="AluSx repeat: matches 1..311 of consensus"
12549..12862
repeat_region
/note="L1MC4 repeat: matches 371..663 of consensus"
12937..13246
repeat_region
/note="AluY repeat: matches 1..309 of consensus"
13294..13389
repeat_region
/note="L1MD2 repeat: matches 6242..6330 of consensus"
13390..13687
repeat_region
/note="AluSx repeat: matches 12..310 of consensus"
13688..13916
repeat_region
/note="L1MD2 repeat: matches 6031..6242 of consensus"
13917..14217
repeat_region
/note="AluSx repeat: matches 1..303 of consensus"
14218..14516
repeat_region
/note="L1MD2 repeat: matches 5660..6031 of consensus"
14522..14691
repeat_region
/note="L1M4 repeat: matches 2357..2493 of consensus"
14692..15025
repeat_region
/note="AluSx repeat: matches 1..310 of consensus"
15026..15055
repeat_region
/note="L1M4 repeat: matches 2324..2357 of consensus"
15158..15588
repeat_region
/note="L1M4 repeat: matches 5725..6183 of consensus"
15605..15784
repeat_region
/note="AluDb repeat: matches 1..271 of consensus"
15785..15850
repeat_region
/note="L1M4 repeat: matches 5667..5733 of consensus"
15851..16168
repeat_region
/note="AluSc repeat: matches 1..306 of consensus"
16169..16739
repeat_region
/note="L1M4 repeat: matches 5176..5667 of consensus"
16740..17052
repeat_region
/note="AluSx repeat: matches 1..312 of consensus"
17053..17519
repeat_region
/note="L1M4 repeat: matches 4696..5176 of consensus"
17541..19073


```

OY 631 GCGTACTTGGCCCTACCAATCTGGTATCCAGTATTCCTCAGTGAATGAGATTCTCG 690
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47341 CCACACACTTGTCCCTGACCTGTTGGATCCATATATCCACATCTTAATTTTSTA 47400

OY 691 CTTCAGTCTACGCAATTCCTCAC-ATAATTCTGACCTACAGAGGAGGATCATATAGC 749
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47401 GTTGTAGTACGTGTGTTCTCTCTGTTAATCTGGATACAGAGCAATCACAGAGC 47460

OY 750 TCTTCAGAGTCTGTGACTCTCCCTCACAAATTCATTCTC-CTGTGTAGTGAAGGTG 808
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47461 TCTTCAGAGGAGGAGGAGTCTCTCCCTCACAAATCTTTCACAAAGTGTGGTAAGGCTA 47520

OY 809 CGCCCTCTGAGCCTCCAGGAGGTGGTGCAGGTC-ACAATGATGATGTATGATCTGTG 867
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47521 TGTCTTCTGAGACCTCCAGCTGGATAGTAGTCTAAAGTACTATTCACACTACCA 47580

OY 868 TTCCCATTAACCAAGCTTT-AAATCCCTCATGCTGCTACACGAGGAGCTACCA 926
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47581 TCCCAATCTCCCTAAGCCTTTGATTTCTCTCTACATTAACCAAGGAGATCAGCA 47640

OY 927 TTT 929
      |||
Db 47641 TTT 47643

RESULT 13
AC005906/c DNA PRI 30-JAN-1999
LOCUS Homo sapiens 12P13.3 BAC RPiC11-429A20 (Roswell Park Cancer
DEFINITION Institute Human BAC Library) complete sequence.
ACCESSION AC005906
NID 94165009
VERSION AC005906.1 GI:4165009
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS Muzny,D., Aronson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y.,
Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H.,
Gorrell,L.L., Hernandez,J., Issar,A., Jackson,L., Kneitz,S.,
Kondejewski,N., Lau,S., Leal,B., Lee,E., Licharge,O., Liu,W.,
Logan,O., Lu,J., Marondel,I., Martinez,C., Merscher,S., Miller,A.,
Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,
Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shum,C., Simon,M.,
Vo,Q., Williamson,A., Worley,K.C., Xiang,A.M., Yang,R., Yu,W.,
Zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A.
Direct Submission
Unpublished
JOURNAL 2 (bases 1 to 185952)
REFERENCE
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 185952)
REFERENCE
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 185952)
REFERENCE
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 185952)
REFERENCE
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

```

COMMENT

On Jan 20, 1999 this sequence version replaced gi:4079596.
 INFORMATION: <http://gc.bcm.tmc.edu:8088/home.html> or email
 gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
 entire insert of this clone. Overlapping regions of clones are only
 sequenced and submitted once, so the sequence for the remainder of
 the insert may be found in the record for the adjacent clones.
 Overlapping clones are noted at the beginning and end of the
 Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches
 of a local database that includes entries from dbSTS, GDB, and
 local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
 unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
 EST and cDNA sequences. Genes demonstrate at least two exons
 flanked by consensus splice sites that maintained sequence
 continuity across the splice junctions. Sequences that are not
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
 standard of double strand coverage with a minimum of 2 clones and 2
 reads with no ambiguities or 2 chemistries with a minimum of 2
 clones and 3 reads with no ambiguities. If the sequence quality for
 a region does not meet this standard, it will be indicated in the
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
 standards - estimated error rate less than 1 per 10,000 bases.
 Reports of lowest quality individual bases and measures of base
 quality are listed below. Description of the metrics can be found
 at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

```

----- Summary Statistics -----
Contig Length: 185961
Phrap values in estimate: 98505
Average error rate (BCM-Phrap estimate): 5.5169e-07
Fraction of Phrap values less than 40 : 0.00108624
Number of consensus changing edits: 4
Number of N's in consensus : 0

```

```

----- Consensus changing edits -----
Position Original+Context Edited+Context
9492 aaagtaac(t)tttttttt aaagtaac(c)tttttttt
36770 ttgggttgggg(n)agcgggaggg ttgggttgggg(g)agcgggaggg
107439 ctggcccttg(n)ctttacact ctggcccttg(a)ctttacact
109251 gctcccgagc(n)ccctactac gctcccgagc(c)ccctactac

```

```

----- Bases with BCM-Phrap value < 20 -----
Quality Position Surrounding Sequence

```

```

----- Distribution of Quality < 40 Bases -----

```

```

1001
901
801
701
601
501
401
301
201
#
bases
*
*
*
*
*
*
*
*

```

101 * * *
01 -----
5 10 15 20 25 30 35 40
Phrap Value Range

Version: 1.01 xfst.

```
FEATURES
Source      Location/Qualifiers
            1..185952 Homo sapiens
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="RPC11-429A20"
            /chromosome="12p13.3"
            2451..2743
repeat_region /rpt_family="L1P"
            2734..2959
repeat_region /rpt_family="L1PA11"
            complement(3054..3118)
repeat_region /rpt_family="L2"
            complement(3332..3426)
repeat_region /rpt_family="MERS3"
            5492..6248
repeat_region /rpt_family="L1ME2"
            6249..6548
repeat_region /rpt_family="AluY"
            6549..6742
repeat_region /rpt_family="L1ME2"
            complement(6743..6825)
repeat_region /rpt_family="(CA)n"
            6847..7184
repeat_region /rpt_family="L1ME2"
            complement(7202..7315)
repeat_region /rpt_family="AluUo"
            7316..7421
repeat_region /rpt_family="L1MA10"
            7951..8224
repeat_region /rpt_family="L1MC/D"
            9327..9418
repeat_region /rpt_family="L1M1"
            complement(9499..9765)
repeat_region /rpt_family="AluY"
            9832..9928
repeat_region /rpt_family="MSTRA"
            10248..10368
repeat_region /rpt_family="MSTRA"
            10374..10668
repeat_region /rpt_family="AluSc"
            10669..10901
repeat_region /rpt_family="MSTRA"
            12777..12960
repeat_region /rpt_family="MIR"
            13055..13459
repeat_region /rpt_family="L1TR20"
            13754..13808
repeat_region /rpt_family="MIR"
            complement(15214..15393)
repeat_region /rpt_family="AluUb"
            15863..16122
repeat_region /rpt_family="L1PB3"
            16203..16361
repeat_region /rpt_family="MIR"
            complement(16372..16554)
repeat_region /rpt_family="MIR"
            complement(16838..16957)
repeat_region /rpt_family="(CA)n"
            complement(22903..23067)
repeat_region /rpt_family="MERSB"
            23484..23972
repeat_region /rpt_family="MLT2FA"
            24144..24174
repeat_region /rpt_family="(CAAAA)n"
```

repeat_region 24212..25128
/rpt_family="MIRVL"
repeat_region complement(25147..25701)
/rpt_family="MLT2D"
repeat_region 25705..25985
/rpt_family="HERVL"
repeat_region 30191..30523
/rpt_family="MLT2E"

Query Match 11.5%; Score 128.8; DB 11; Length 185952;
Best Local Similarity 68.0%; Pred. No. 1e-25;
Matches 227; Conservative 0; Mismatches 97; Indels 10; Gaps 3;

```
QY 518 GAACACTGTGATCACTACACGACCATATGTTCCGACAGACTATCATGATTAAT 577
      |||||
DB 26093 GAACACTGTGATCACTACACGACCATATGTTCCGACAGACTATCATGATTAAT 26034
      |||||

QY 578 GTGTTGACTGTGCTGTCTATTTGTTACTATACCATGCGGA---TGTTAGTGAAATTAAGCGT 634
      |||||
DB 26033 GCTTTGACTTTGTTGTCATTAAGTAACACACCAATTAATTTGTTGTTAAGTGCTGT 25974
      |||||

QY 635 CACTTGCCCTCAACCATCTTGATTCACATTAATCTCATGATTAATGAGATTCCGCTTC 694
      |||||
DB 25973 CACTTGCCCTCAACCATCTTGATTCACATTAATCTCATGATTAATGAGATTCCAGTTC 25914
      |||||

QY 695 AGTGTGACCATTCCTCA-CATTAATTTCTGACCTACAGAGTGAGGATCATATAGCTCTT 753
      |||||
DB 25913 AGTGTGACCATTCCTCAACCATATTTCTCACTGTAGAGAGAGACATACAGTCTCTT 25854
      |||||

QY 754 CAAGATGCTGTGATCCCTCCACAAATTCATTTCTCTGTTGTAGTGAAGTGCGGCC 813
      |||||
DB 25853 CATGATGCTGTGAGACTCCCTCAAAATTTATTTTC-----TTTAGTGAAGATGTGTC 25800
      |||||

QY 814 TCTGAGCCTCCACGAGTGCTGTGAGTGCAC 847
      |||||
DB 25799 TTTGATCTTCTTGCTGTGAGTGCAC 25766
      |||||
```

RESULT 14

HS44N10/c

LOCUS HS44N10 208643 bp DNA HTG 12-NOV-1998
DEFINITION Homo sapiens chromosome 12 clone 44N10, WORKING DRAFT SEQUENCE, in
unordered pieces.

ACCESSION

93873505

VERSION

297197.1 GI:3873505

KEYWORDS

HTG; HTGS_PHASE1.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 208643)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

White, S.
Submitted (12-NOV-1998) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humanyesanger.ac.uk Clone requests: clonequest@esanger.ac.uk
On Nov 16, 1998 this sequence version replaced gi:12225928.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E. coli, yeast, vector,
phase etc. Order of segments is not known; 800 n.s. separate
segments. Unfinished: 44N10 Contig-ID: 03188 Length: 28250 bp
Unfinished: 44N10 Contig-ID: 03967 Length: 7858 bp Unfinished:
44N10 Contig-ID: 04550 Length: 16612 bp Unfinished: 44N10
Contig-ID: 03869 Length: 1065 bp Unfinished: 44N10 Contig-ID: 00878
Length: 24110 bp Unfinished: 44N10 Contig-ID: 04458 Length: 26725
bp Unfinished: 44N10 Contig-ID: 04588 Length: 6852 bp Unfinished:
44N10 Contig-ID: 04631 Length: 9284 bp Unfinished: 44N10 Contig-ID:
02809 Length: 1644 bp Unfinished: 44N10 Contig-ID: 04439 Length:
1307 bp Unfinished: 44N10 Contig-ID: 04403 Length: 1278 bp

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

Search completed: September 28, 1999, 12:39:06
Job time: 6026 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 12:28:42 : Search time 289.74 Seconds
(without alignments)
966.263 Million cell updates/sec

Title: US-09-030-606-177

Perfect score: 1119

Sequence: 1 GCGCAGCTCGCAGCCCTGCA.....ATGTTAAAAA 1119

Scoring table: IDENTITY_NUC

Searched: 311565 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1119	100.0	1119	1	V58648
2	1119	100.0	1119	1	V61253
3	361.4	32.3	1265	1	V58645
4	361.4	32.3	1265	1	V61250
5	352.2	31.5	1167	1	V58647
6	352.2	31.5	1167	1	V61252
7	350.4	31.3	871	1	V37495
8	339.8	30.4	1386	1	V11855
9	338.8	30.3	1248	1	V58644
10	338.8	30.3	1248	1	V61249
11	338	30.2	402	1	X41114
12	257.4	23.0	1459	1	V58646
13	257.4	23.0	1459	1	V61251
14	153.8	13.7	234	1	V58522
15	153.8	13.7	234	1	V61168
16	100.4	9.0	1476	1	X16285
17	93	8.3	907	1	N70905
18	91	8.1	558	1	O20557
19	91	8.1	379	1	O20555
20	91	8.1	555	1	O20556
21	91	8.1	309	1	O20558
22	91	8.1	309	1	O20559
23	89.8	8.0	5406	1	N93197
24	85	7.6	738	1	N71049
25	81	7.2	986	1	O81203
26	81	7.2	1089	1	T39783
27	79	7.1	450	1	O20560
28	76.4	6.8	832	1	T05147
29	76.4	6.8	832	1	V06609
30	76.4	6.8	832	1	V06604
31	76.4	6.8	832	1	V32938
32	76.4	6.8	832	1	V70340
33	76.4	6.8	832	1	X08946
34	76.2	6.8	798	1	T46175
35	75.2	6.7	925	1	N93196
36	73.4	6.6	6139	1	V70354
37	73.2	6.5	1466	1	V32496
38	72.8	6.5	992	1	T91055
39	72.6	6.5	1462	1	T91055
40	72.2	6.4	200	1	V60379
41	71.6	6.4	766	1	T05149
42	71.6	6.4	766	1	V06603
43	71.6	6.4	766	1	V70341

ALIGNMENTS

RESULT	ID	Score	Query Match	Length	DB ID	Description
1	V58648	1119	100.0	1119	1	V58648
2	V58648	1119	100.0	1119	1	V58648
3	V58648	1119	100.0	1119	1	V58648
4	V58648	1119	100.0	1119	1	V58648
5	V58648	1119	100.0	1119	1	V58648
6	V58648	1119	100.0	1119	1	V58648
7	V58648	1119	100.0	1119	1	V58648
8	V58648	1119	100.0	1119	1	V58648
9	V58648	1119	100.0	1119	1	V58648
10	V58648	1119	100.0	1119	1	V58648
11	V58648	1119	100.0	1119	1	V58648
12	V58648	1119	100.0	1119	1	V58648
13	V58648	1119	100.0	1119	1	V58648
14	V58648	1119	100.0	1119	1	V58648
15	V58648	1119	100.0	1119	1	V58648
16	V58648	1119	100.0	1119	1	V58648
17	V58648	1119	100.0	1119	1	V58648
18	V58648	1119	100.0	1119	1	V58648
19	V58648	1119	100.0	1119	1	V58648
20	V58648	1119	100.0	1119	1	V58648
21	V58648	1119	100.0	1119	1	V58648
22	V58648	1119	100.0	1119	1	V58648
23	V58648	1119	100.0	1119	1	V58648
24	V58648	1119	100.0	1119	1	V58648
25	V58648	1119	100.0	1119	1	V58648
26	V58648	1119	100.0	1119	1	V58648
27	V58648	1119	100.0	1119	1	V58648
28	V58648	1119	100.0	1119	1	V58648
29	V58648	1119	100.0	1119	1	V58648
30	V58648	1119	100.0	1119	1	V58648
31	V58648	1119	100.0	1119	1	V58648
32	V58648	1119	100.0	1119	1	V58648
33	V58648	1119	100.0	1119	1	V58648
34	V58648	1119	100.0	1119	1	V58648
35	V58648	1119	100.0	1119	1	V58648
36	V58648	1119	100.0	1119	1	V58648
37	V58648	1119	100.0	1119	1	V58648
38	V58648	1119	100.0	1119	1	V58648
39	V58648	1119	100.0	1119	1	V58648
40	V58648	1119	100.0	1119	1	V58648
41	V58648	1119	100.0	1119	1	V58648
42	V58648	1119	100.0	1119	1	V58648
43	V58648	1119	100.0	1119	1	V58648

|||||
Db 361 GATGCTGTGATGGCATCCAGTCCCAAGACTGTGGAGGCTGGGAGTGTGAAGCTTTC 420
421 CAACCCGAGAGGTTTACCATTTCCGCAACTTCCAGTGCAGAGAGTCTCTGTGATC 480
421 CAACCCGAGAGGTTTACCATTTCCGCAACTTCCAGTGCAGAGAGTCTCTGTGATC 480
481 CTGACTGGGTGCTCACTACTGCTCACTGCATCAACCCGAGACATGATCACTAGCCAG 540
481 CTGACTGGGTGCTCACTACTGCTCACTGCATCAACCCGAGACATGATCACTAGCCAG 540
541 CACCATAGTTCTCCGAAGTCAAGCATATCATATGATCTGTGTGACTGTCTATTTGT 600
541 CACCATAGTTCTCCGAAGTCAAGCATATCATATGATCTGTGTGACTGTCTATTTGT 600
601 ACTAACCATCCCATGTTTATGAGTAATAGCCATCTTGGCTCAACCATTTGGATC 660
601 ACTAACCATCCCATGTTTATGAGTAATAGCCATCTTGGCTCAACCATTTGGATC 660
661 CAGTTATCCCTCACTGAATGAGATTTCTGCTTCAAGTGTGAGCCATCCCAATATTTTC 720
661 CAGTTATCCCTCACTGAATGAGATTTCTGCTTCAAGTGTGAGCCATCCCAATATTTTC 720
721 TGAACCTACAGAGGTGAGGATCATATATAGCTCTTCAAGGATGCTGTACTCCCTCACAAA 780
721 TGAACCTACAGAGGTGAGGATCATATATAGCTCTTCAAGGATGCTGTACTCCCTCACAAA 780
781 TTGATTTCTCTCTGTGTAGTAAAGGTGCGCCCTCTGAGCCCTCCAGGGTGGGTGCA 840
781 TTGATTTCTCTCTGTGTAGTAAAGGTGCGCCCTCTGAGCCCTCCAGGGTGGGTGCA 840
841 GGTCAATGATGAATGATGATGCTGTCTCCATTAACCAAGCCTTTAAATCCCTATG 900
841 GGTCAATGATGAATGATGATGCTGTCTCCATTAACCAAGCCTTTAAATCCCTATG 900
901 CTGAGTACACCAAGGAGGCTGTACATTTCTCTATTAGTGTATGCTGCCATTCATGCA 960
901 CTGAGTACACCAAGGAGGCTGTACATTTCTCTATTAGTGTATGCTGCCATTCATGCA 960
961 ACCACCTCAGAGACTCGAGATCTCTGCTAGTGAAGTCTGCTGATCTGCTGCTTGGG 1020
961 ACCACCTCAGAGACTCGAGATCTCTGCTAGTGAAGTCTGCTGATCTGCTGCTTGGG 1020
1021 GAGGTGAGGAGAGAGGCTGATGTTCAATGGATCTGTGAGTGTAAACATATAGTGC 1080
1021 GAGGTGAGGAGAGAGGCTGATGTTCAATGGATCTGTGAGTGTAAACATATAGTGC 1080
1081 TTATTAACAGAGCTGTGATGTTAAAAA 1119
1081 TTATTAACAGAGCTGTGATGTTAAAAA 1119

RESULT 2
V61253
ID V61253 standard; cDNA; 1119 BP.
AC V61253;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE14.
KM Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN W09837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; 003492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillion DC, Xu J;
DR WPI: 98-609886/51.
DR P-PDB: W71873.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 108-109; 130pp; English.

CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match 100.0%; Score 1119; DB 1; Length 1119;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GCCCAGCTGAGAGCCCTGCGAGGCGGACATGTCATGAAACAAATGTTCTGCTGGGG 60
1 GCCCAGCTGAGAGCCCTGCGAGGCGGACATGTCATGAAACAAATGTTCTGCTGGGG 60
61 GTCCTGGTGCATCCGAGTGGGTGCTGTACAGCCGACAGTGTTCAGAACTCTACACC 120
61 GTCCTGGTGCATCCGAGTGGGTGCTGTACAGCCGACAGTGTTCAGAACTCTACACC 120
121 ATCGGCTGGGCTGCAAGTCTTTGAGCGCCGACCAAGAGCCAGAGTGTGAG 180
121 ATCGGCTGGGCTGCAAGTCTTTGAGCGCCGACCAAGAGCCAGAGTGTGAG 180
181 GCCAGCCTCCGTTAGGAGCCAGAGTACACAGACCCCTGTGCTAACGACCTCATG 240
181 GCCAGCCTCCGTTAGGAGCCAGAGTACACAGACCCCTGTGCTAACGACCTCATG 240
241 CTGATCAAGTTGAGCAAGATCCGTCGAGTGTGACACCATCCGAGCATCAATTTGCT 300
241 CTGATCAAGTTGAGCAAGATCCGTCGAGTGTGACACCATCCGAGCATCAATTTGCT 300
301 TCCGAGTGGCTTACCGCGGGGAACTCTGCTGTTTGTGCTGGGGTGTGCTGGGAA 360
301 TCCGAGTGGCTTACCGCGGGGAACTCTGCTGTTTGTGCTGGGGTGTGCTGGGAA 360
361 GATGCTGTGATTCATCCATCCAGCTGTCGAGAGGCTGGGAGTGTGAAGCTTTCC 420
361 GATGCTGTGATTCATCCATCCAGCTGTCGAGAGGCTGGGAGTGTGAAGCTTTCC 420
421 CAACCCGAGAGGTTGTACATTTGCGCAACTTCCAGTGCAGAGAGTCTCTGCATC 480
421 CAACCCGAGAGGTTGTACATTTGCGCAACTTCCAGTGCAGAGAGTCTCTGCATC 480
481 CTGACTGGGTGCTCACTACTGCTCACTGCATCAACCCGAGACATGATCACTAGCCAG 540
481 CTGACTGGGTGCTCACTACTGCTCACTGCATCAACCCGAGACATGATCACTAGCCAG 540
541 CACCATAGTTCTCCGAAGTCAAGTATGATGATGATGATGATGATGATGATGATGAT 600
541 CACCATAGTTCTCCGAAGTCAAGTATGATGATGATGATGATGATGATGATGATGAT 600
601 ACTAACCATCCCATGTTTATGAGTAATAGCCATCTTGGCTCAACCATTTGGATC 660
601 ACTAACCATCCCATGTTTATGAGTAATAGCCATCTTGGCTCAACCATTTGGATC 660
661 CAGTTATCCCTCACTGAATGAGATTTCTGCTTCAAGTGTGAGCCATCCCAATATTTTC 720
661 CAGTTATCCCTCACTGAATGAGATTTCTGCTTCAAGTGTGAGCCATCCCAATATTTTC 720
721 TGAACCTACAGAGGTGAGGATCATATATAGCTCTTCAAGGATGCTGTACTCCCTCACAAA 780
721 TGAACCTACAGAGGTGAGGATCATATATAGCTCTTCAAGGATGCTGTACTCCCTCACAAA 780
781 TTGATTTCTCTCTGTGTAGTAAAGGTGCGCCCTCTGAGCCCTCCAGGGTGGGTGCA 840
781 TTGATTTCTCTCTGTGTAGTAAAGGTGCGCCCTCTGAGCCCTCCAGGGTGGGTGCA 840
841 GGTCAATGATGAATGATGATGCTGTCTCCATTAACCAAGCCTTTAAATCCCTATG 900
841 GGTCAATGATGAATGATGATGCTGTCTCCATTAACCAAGCCTTTAAATCCCTATG 900

OY	901	CACAGTACGAGGGAGGTCTAGCATTTCTTCATTATAGTAGTGCCATATGCA	960
OY	901	CGCAGTACGAGGGAGGTCTAGCATTTCTTCATTATAGTAGTGCCATATGCA	960
DB	901	CTCAGTACGCCAGGCGAGGTCTAGCATTTCTTCATTATAGTAGTGCCATATGCA	960
OY	961	ACCACCTCAGGACTCCTGGATTCTCTGCAGTTGAGCTCTGCATGCTCCCTCTGGG	1020
DB	961	ACCACCTCAGGACTCCTGGATTCTCTGCAGTTGAGCTCTGCATGCTCCCTCTGGG	1020
OY	1021	GAGGTGAGGGAAGAGGCCCATGGTTCAATGAGATCTGTGCAAGTTGTAACAATGAGTGC	1080
DB	1021	GAGGTGAGGGAAGAGGCCCATGGTTCAATGAGATCTGTGCAAGTTGTAACAATGAGTGC	1080
OY	1081	TAAATAAACAGAAGCGTGATGTAAAAA1119	
DB	1081	TAAATAAACAGAAGCGTGATGTAAAAA1119	
 RESULT 3 V58645 V58645 standard; cDNA: 1265 BP.			
DT	V58645:		
DE	08-DEC-1998 (first entry)		
KW	Prostate tumour specific gene clone DE2.		
OS	Prostate tumour specific gene; human; prostate cancer; detection;		
KS	therapy; ss.		
PN	Homo sapiens.		
PF	W09837418-A2.		
PR	27-AUG-1998.		
PR	25-FEB-1998; U03690.		
PR	09-FEB-1998; US-904809.		
PR	25-FEB-1997; US-806586.		
PA	01-AUG-1997; US-904809.		
PI	(CORI-) CORIXA CORP.		
DR	Dillon DC, Xu J;		
PT	WPI; 98-480805/41.		
PT	Novel human prostatic specific tumour protein and fragments - useful		
PT	for detecting and treating prostate cancers		
PS	Claim 1; Page 113-114; 14pp; English.		
CC	This sequence represents a human prostate tumour specific gene, and can		
CC	be used in the method of the invention. The method is for detecting		
CC	prostate cancer comprises contacting a biological sample with an agent		
CC	able to bind an immunogenic portion of a prostate protein (such as		
CC	encoded by this sequence). An antibody which binds to an immunogenic		
CC	portion of the prostate protein, and the method can be used to detect,		
CC	monitor progression of, or treat prostate cancers. The antibody may		
CC	also be conjugated to a therapeutic agent for use in therapy of prostate		
CC	cancers.		
SQ	Sequence 1265 BP; 256 A; 432 C; 321 G; 245 T;		
 Query Match 32.3%; Score 361.4; DB 1; Length 1265; Best Local Similarity 99.7%; Pred. No. 6.5e-98;			
Matches	362; Conservative	0; Mismatches	1; Indels 0; Gaps 0;
OY	2	CGCACTCGAGCCCTGCGAGCGCGGCACTGGTCATGGAAAACGAATGTTCTGCTGGGCG	61
DB	8	CGCACTCGAGCCCTGCGAGCGCGGCACTGGTCATGGAAAACGAATGTTCTGCTGGGCG	67
OY	62	TCCGTGTGATCCGAGTGGGTGCTGTCTAGCCGGCANACTGTTCCGAGACTCCTACACA	121
DB	68	TCCGTGTGATCCGAGTGGGTGCTGTCTAGCCGGCANACTGTTCCGAGACTCCTACACA	127
OY	122	TGGGCTGGGCTGCGACATCTTGAAGCGCGACCAAGAGCGAGGAGCCAGATGTGGAG	181
DB	128	TGGGCTGGGCTGCGACATCTTGAAGCGCGACCAAGAGCGAGGAGCCAGATGTGGAG	187
OY	182	CCAGCCTCTCGGTAGGCGCACCCAGATACACAGACCCCTTGTCTGCTAAGACCTCATGC	241
DB	188	CCAGCCTCTCGGTAGGCGCACCCAGATACACAGACCCCTTGTCTGCTAAGACCTCATGC	247
OY	242	TGATCAAGTTGGAACGAATCCGTTGCCGATCTGACACCAATCCGGAGCATCAGATTGCTT	301
DB	248	TGATCAAGTTGGAACGAATCCGTTGCCGATCTGACACCAATCCGGAGCATCAGATTGCTT	307

Query Match	Best Local Similarity	32.3%	Score 361.4	DB 1	Length 1265
Matches 362	Conservative	0	Mismatches 1	Indels	Gaps
OY	2	CGCACTCCAGCCCGGAGGCGGACATGTCATGGAAGAATGTTCTGCTGGGCG	61		
Db	8	CGCACTCCAGCCCGGAGGCGGACATGTCATGGAAGAATGTTCTGCTGGGCG	67		
OY	62	TCCTGTCATCCGAGTGGTGTCGTGACGCGACACTGTTCCAGAACTCTACACCA	121		
Db	68	TCCTGTCATCCGAGTGGTGTCGTGACGCGACACTGTTCCAGAACTCTACACCA	127		
OY	122	TCGGGCTGGGCTGACACAGTCTTAGGCGCCAGCAAGAGCCAGGAGCCAGATGGTGGAG	181		
Db	128	TCGGGCTGGGCTGACACAGTCTTAGGCGCCAGCAAGAGCCAGGAGCCAGATGGTGGAG	187		
OY	182	CCAGGCTCTCCGTAGGAGCCAGACAGATCAACAGACCTTGGTCTCTAAGACCTCATCC	241		
Db	188	CCAGGCTCTCCGTAGGAGCCAGACAGATCAACAGACCTTGGTCTCTAAGACCTCATCC	247		
OY	242	TCATCAAGTTGGAGCAATCCGTGTCGAGTGTGACACCATCCGAGACATCAGATGCTT	301		
Db	248	TCATCAAGTTGGAGCAATCCGTGTCGAGTGTGACACCATCCGAGACATCAGATGCTT	307		
OY	302	CGCATGCGCCATACCGCGGGGGAACCTTCCCTTCCTTGGCGGGGCTGCTGGGGAACG	361		
Db	308	CGCATGCGCCATACCGCGGGGGAACCTTCCCTTCCTTGGCGGGGCTGCTGGGGAACG	367		
OY	362	ATG 364			
Db	368	GTG 370			

V58647
ID V58647 standard; cDNA; 1167 BP.
AC V58647;
DE 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE13.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 28..645
ET /tag= a
FN WO9837418-A2.
PD 27-AUG-1998.
PR 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
DR P-PSDB; W69388.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 115; 141p; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprising contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancer.
SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;

Query Match 31.5%; Score 352.2; DB 1; Length 1167;
Best Local Similarity 96.3%; Pred. No. 3.4e-95;
Matches 360; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 8 CGCAGCCCTGGGAGGCGGACATGTCATGGAAGAAAGCAATTGTTGCTGGGCGTCTGG 67
DB 2 CGCAGCCCTGGGAGGCGGACATGTCATGGAAGAAAGCAATTGTTGCTGGGCGTCTGG 61
QY 68 TGCATCCGAGTGGGTGCTGTGACGCGACACTGTTCCAGAACTCTACACATCGGCG 127
DB 62 TGCATCCGAGTGGGTGCTGTGACGCGACACTGTTCCAGAACTCTACACATCGGCG 121
QY 128 TGGGCTGACAGTCTTGAAGCGGACCAAGAGCCAGAGGAGGAGATGTGAGGCGCAGC 187
DB 122 TGGGCTGACAGTCTTGAAGCGGACCAAGAGCCAGAGGAGGAGATGTGAGGCGCAGC 181
QY 188 TCTCCGTAGGAGCAACCAAGTACAAAGACCTGTGCTTACAGCACTCATGCTATCA 247
DB 182 TCTCCGTAGGAGCAACCAAGTACAAAGACCTGTGCTTACAGCACTCATGCTATCA 241
QY 248 AGTTGGAGCAATCCGTGCTGAGTGTGACACCATCCGAGCAATCCATTCGTCAGT 307
DB 242 AGTTGGAGCAATCCGTGCTGAGTGTGACACCATCCGAGCAATCCATTCGTCAGT 301
QY 308 GCCCTACCGCGGGGAACTCTTGGCTTGGCTGGGCTGCTGCGGAAAGATGCTG 367
DB 302 GCCCTACCGCGGGGAACTCTTGGCTTGGCTGGGCTGCTGCGGAAAGATGCTG 361
QY 368 TGATTGCCATCCAG 381
DB 362 TGCTACCGTGTCTG 375

RESULT 6
ID V61252
AC V61252 standard; cDNA; 1167 BP.

DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE13.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 28..645
ET /tag= a
FN WO9837093-A2.
PD 27-AUG-1998.
PR 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
DR P-PSDB; W71782.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 107; 130p; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;

Query Match 31.5%; Score 352.2; DB 1; Length 1167;
Best Local Similarity 96.3%; Pred. No. 3.4e-95;
Matches 360; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 8 CGCAGCCCTGGGAGGCGGACATGTCATGGAAGAAAGCAATTGTTGCTGGGCGTCTGG 67
DB 2 CGCAGCCCTGGGAGGCGGACATGTCATGGAAGAAAGCAATTGTTGCTGGGCGTCTGG 61
QY 68 TGCATCCGAGTGGGTGCTGTGACGCGACACTGTTCCAGAACTCTACACATCGGCG 127
DB 62 TGCATCCGAGTGGGTGCTGTGACGCGACACTGTTCCAGAACTCTACACATCGGCG 121
QY 128 TGGGCTGACAGTCTTGAAGCGGACCAAGAGCCAGAGGAGGAGATGTGAGGCGCAGC 187
DB 122 TGGGCTGACAGTCTTGAAGCGGACCAAGAGCCAGAGGAGGAGATGTGAGGCGCAGC 181
QY 188 TCTCCGTAGGAGCAACCAAGTACAAAGACCTGTGCTTACAGCACTCATGCTATCA 247
DB 182 TCTCCGTAGGAGCAACCAAGTACAAAGACCTGTGCTTACAGCACTCATGCTATCA 241
QY 248 AGTTGGAGCAATCCGTGCTGAGTGTGACACCATCCGAGCAATCCATTCGTCAGT 307
DB 242 AGTTGGAGCAATCCGTGCTGAGTGTGACACCATCCGAGCAATCCATTCGTCAGT 301
QY 308 GCCCTACCGCGGGGAACTCTTGGCTTGGCTGGGCTGCTGCGGAAAGATGCTG 367
DB 302 GCCCTACCGCGGGGAACTCTTGGCTTGGCTGGGCTGCTGCGGAAAGATGCTG 361
QY 368 TGATTGCCATCCAG 381
DB 362 TGCTACCGTGTCTG 375

RESULT 7
ID V37495
AC V37495 standard; DNA; 871 BP.
DE 07-SEP-1998 (first entry)
DE Human prostate-specific kallikrein (HPSK) encoding DNA.
KW Prostate-specific kallikrein; HPSK; prostate carcinoma; human;
KW benign prostate hyperplasia; diagnosis; drug screening; PSK; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 31..777
ET /tag= a
ET /transl_except= (pos:367..369, aa:Xaa)
ET /transl_except= (pos:412..414, aa:Xaa)

FT /transl_except- (pos:424, .426, aa:Xaa)
 FT /product- "HPSK protein"
 FT /note- "Xaa - unknown"
 PN W09820117-A1.
 PD 14-MAY-1998.
 PF 31-OCT-1997; U20051.
 PR 05-NOV-1996; US-744026.
 PA (INCY-) INCYTE PHARM INC.
 PI Bandman O, Goli SK.
 DR WPI: 98-286933/25.
 P-PSDB; W60592.
 PT New isolated prostate-specific kallikrein - used to develop products
 PT for diagnosis and treatment of, e.g. prostate carcinoma or benign
 PT hyperplasia
 PS Claim 5: Fig 1A-C; 68pp; English.
 CC This DNA encodes a human prostate-specific kallikrein (HPSK). A host cell
 CC containing an expression vector comprising the HPSK nucleic acid sequence
 CC can be used to produce the protein recombinantly. The HPSK products can
 CC be used for the diagnosis of conditions or diseases associated with
 CC expression of HPSK such as prostate carcinoma and benign prostate
 CC hyperplasia. Agonists and antagonists which specifically bind to HPSK and
 CC modulate its activity can be used for the preparation of treatment of
 CC such conditions or diseases. The products can also be used for detection
 CC and drug screening, especially for the detection of prostate-specific
 CC kallikrein (PSK).
 SQ Sequence 871 BP; 166 A; 260 C; 258 G; 184 T;

Query Match 31.3%; Score 350.4; DB 1; Length 871;
 Best Local Similarity 95.0%; Pred. No. 1e-94;
 Matches 360; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 3 GCATCCGAGCCCTGGCGGCGGCGCATGTGTAAGAAAGCAATGTTCTGCGGGCT 62
 DB 129 GCATCCGAGCCCTGGCGGCGGCGCATGTGTAAGAAAGCAATGTTCTGCGGGCT 188
 QY 63 CCTGGTGCATCCGAGTGGTGTCTGTACCGCAGCACTGTTTCCAGAACTCCTACACAT 122
 DB 189 CCTGGTGCATCCGAGTGGTGTCTGTACCGCAGCACTGTTTCCAGAACTCCTACACAT 248
 QY 123 CGGAGCTGGGCTGCACACTTGTAGGCGGACCAAGCCAGGAGCCAGATGTTGAGAGC 182
 DB 249 CGGAGCTGGGCTGCACACTTGTAGGCGGACCAAGCCAGGAGCCAGATGTTGAGAGC 308
 QY 183 CAGGCTCTCCCTAGGCGGCGGCGCATGTGTAAGAAAGCAATGTTCTGCGGGCT 242
 DB 309 CAGGCTCTCCCTAGGCGGCGGCGCATGTGTAAGAAAGCAATGTTCTGCGGGCT 368
 QY 243 CATCAAGTTGAGCAATCCGTGTCCAGTGTGACACATCCGAGACATCAGATGTTT 302
 DB 369 CATCAAGTTGAGCAATCCGTGTCCAGTGTGACACATCCGAGACATCAGATGTTT 428
 QY 303 GCAATGCCCTACCGGCGGGAATCTTGGCTGTGCTGGGGTGTCTGGCGGACGA 362
 DB 429 GCAATGCCCTACCGGCGGGAATCTTGGCTGTGCTGGGGTGTCTGGCGGACGA 488
 QY 363 TGCCTGTATTTGCCATCCAG 381
 DB 489 CAGATGCTCTACCGTGTG 507

RESULT 8
 V11855
 ID V11855 standard; cDNA; 1386 BP.
 AC V11855;
 DT 11-SEP-1998 (first entry)
 DE Homo sapiens Tub Interactor (hTI-1) gene.
 KW serine protease; tub interactor; treatment; obesity; cachexia;
 KW anorexia nervosa; diabetes; cell cycle progression; apoptosis;
 KW neurodegenerative disease; Alzheimer's disease; drug screening;
 KW Parkinson's disease; Huntington's chorea; detection; diagnosis;
 KW amyotrophic lateral sclerosis; spinocerebellar degeneration; ss.
 OS Homo sapiens.

EH Key Location/Qualifiers
 FT CDS 2..701
 FT /*tag- a
 FT /product- hti-1 protein
 FT /note- "putative serine protease"
 PN W09812302-A1.
 PD 26-MAR-1998.
 PF 05-SEP-1997; U15627.
 PR 21-JUL-1997; US-897340.
 PR 17-SEP-1996; US-715032.
 PA (MILL-) MILLENNIUM PHARM INC.
 PI Extrada PR, Gimeno CJ.
 DR WPI: 98-217246/19.
 P-PSDB; W59129.
 PT Tub interactor genes - used to develop products for the treatment
 PT of obesity, cachexia, anorexia nervosa or related disorders e.g.
 PT diabetes
 PS Claim 10: Fig 1; 120pp; English.
 CC The sequence is that of the Tub Interactor gene hti-1 which
 CC codes for a putative serine protease. TI genes function
 CC in biochemical pathways involved in weight control and
 CC related disorders. The products can be used for treating
 CC weight disorders, e.g. obesity, cachexia or anorexia nervosa,
 CC or a related disorder such as diabetes. The products can
 CC also be used to modulate cell cycle progression and apoptosis.
 CC They can be used for treating neurodegenerative diseases
 CC which are characterised by apoptosis, including Alzheimer's
 CC disease, Parkinson's disease, Huntington's chorea, amyotrophic
 CC lateral sclerosis or spinocerebellar degenerations. The
 CC products can also be used for detection, diagnosis and
 CC drug screening.
 SQ Sequence 1386 BP; 318 A; 490 C; 321 G; 249 T;

Query Match 30.4%; Score 339.8; DB 1; Length 1386;
 Best Local Similarity 93.9%; Pred. No. 1.8e-91;
 Matches 369; Conservative 0; Mismatches 12; Indels 12; Gaps 1;

QY 1 GCGAGCTGCGAGCCCTGGCGGCGGCGCATGTGTAAGAAAGCAATGTTCTGCTCGGGC 60
 DB 14 GCGAGCTGCGAGCCCTGGCGGCGGCGCATGTGTAAGAAAGCAATGTTCTGCTCGGGC 73
 QY 61 GTCTGTGTGATCCGAGTGGTGTCTGTACGCCAGCACTGTTTCCAGAA----- 111
 DB 74 GTCTGTGTGATCCGAGTGGTGTCTGTACGCCAGCACTGTTTCCAGAAATGATGAG 133
 QY 111 --GTCTTACCACTGGGCTGGGCTGTGCACAGTGTGAGGCCGACAGAGCCAGGGAGC 168
 DB 134 AGCTCTTACCACTGGGCTGGGCTGTGCACAGTGTGAGGCCGACAGAGCCAGGGAGC 193
 QY 169 CAGATGTGGAGGCGGCGGCGGCGCATGTGTAAGAAAGCAATGTTCTGCTCGGGC 228
 DB 194 CAGATGTGGAGGCGGCGGCGGCGCATGTGTAAGAAAGCAATGTTCTGCTCGGGC 253
 QY 229 AAGCACTCATGCTCATCAATGTTGAGCAATCCGTGTCCAGTGTGACACATCCGAGC 288
 DB 254 AAGCACTCATGCTCATCAATGTTGAGCAATCCGTGTCCAGTGTGACACATCCGAGC 313
 QY 289 ATCAGCATGCTTGTGACAGTGGCTTACCGGCGGGAATCTTGGCTGTGCTGGGGT 348
 DB 314 ATCAGCATGCTTGTGACAGTGGCTTACCGGCGGGAATCTTGGCTGTGCTGGGGT 373
 QY 349 CTGCTGGCGAAGCATGCTGTGATGCGCATCCAG 381
 DB 374 CTGCTGGCGAAGCATGCTGTGATGCGCATCCGCTGCTG 406

RESULT 9
 V58644
 ID V58644 standard; cDNA; 1248 BP.
 AC V58644;
 DT 08-DEC-1998 (first entry)
 DE Prostate tumour specific gene clone DEL.

KW Prostate tumour specific gene; human; prostate cancer; detection;
 KM therapy; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 217..696
 FT /tag= a
 PD WO9837418-A2.
 PD 27-AUG-1998.
 PF 25-FEB-1998: U03690.
 PR 09-FEB-1998: US-904809.
 PR 25-FEB-1997: US-806596.
 PR 01-AUG-1997: US-904809.
 PA (CORI-) CORIXA CORP.
 PI Dillon DC, Xu J;
 DR WPI: 98-480805/41.
 DR P-PSDB: W69387.
 PT Novel human prostate specific tumour protein and fragments - useful
 PT for detecting and treating prostate cancers
 PS Claim 1: Page 112: 141pp; English.
 CC This sequence represents a human prostate tumour specific gene, and can
 CC be used in the method of the invention. The method is for detecting
 CC prostate cancer comprises contacting a biological sample with an agent
 CC able to bind an immunogenic portion of a prostate protein (such as
 CC encoded by this sequence). An antibody which binds to an immunogenic
 CC portion of the prostate protein, and the method can be used to detect,
 CC monitor progression of, or treat prostate cancers. The antibody may
 CC also be conjugated to a therapeutic agent for use in therapy of prostate
 CC cancers.
 SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 30.3%; Score 338.8; DB 1; Length 1248;
 Best Local Similarity 93.9%; Pred. No. 3.4e-91;
 Matches 368; Conservative 0; Mismatches 12; Indels 12; Gaps 1;

QY 2 CGCAGCTGCGAGCCCTGGCAGGCGGCGACTGCTATGTAAGAAAGCAATGTTCTGCTGGGCG 61
 DB 35 CGCAGCTGCGAGCCCTGGCAGGCGGCGACTGCTATGTAAGAAAGCAATGTTCTGCTGGGCG 94
 QY 62 TCCGTGGTGCATCCGAGTGGGTGCTGTCAAGCCGACACTGTTTCCAGAA----- 111
 DB 95 TCCGTGGTGCATCCGAGTGGGTGCTGTCAAGCCGACACTGTTTCCAGAAAGTGAAGTGCAGA 154
 QY 111 -CTCTACACCATCGGCGGCTGCGTGCAGAGTCTTGAAGCGGACCAAGAGCGAGAGCC 169
 DB 155 GCTCTACACCATCGGCGGCTGCGTGCAGAGTCTTGAAGCGGACCAAGAGCGAGAGCC 214
 QY 170 AGATGTGAGAGCGCAGCTCTCGCTACGCGACCCAGATACAGACAGACCTTGTGCTGA 229
 DB 215 AGATGTGAGAGCGCAGCTCTCGCTACGCGACCCAGATACAGACAGACCTTGTGCTGA 274
 QY 230 ACGACCTCATGCTCATCAAGTTGAGCAATCCGTGCGAGTCTGACACCATCCGGAGCA 289
 DB 275 ACGACCTCATGCTCATCAAGTTGAGCAATCCGTGCGAGTCTGACACCATCCGGAGCA 334
 QY 290 TCAGCATGCTTTCAGAGTGCCTTACCGCGGGAACCTTGGCTGCTTGGTGGGCTC 349
 DB 335 TCAGCATGCTTTCAGAGTGCCTTACCGCGGGAACCTTGGCTGCTTGGTGGGCTC 394
 QY 350 TGGTGGCGAAGCATGCTGTGATTCATCCATCCAG 381
 DB 395 TGGTGGCGAAGCATGCTGTGATTCATCCATCCAG 426

RESULT 10
 ID V61249 standard; cDNA; 1248 BP.
 AC V61249;
 DT 06-JAN-1999 (first entry)
 DE cDNA sequence of prostate tumour clone P703 splice variant DEL.
 KM Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
 OS Homo sapiens.
 PN WO9837093-A2.

PD 27-AUG-1998.
 PF 25-FEB-1998: U03492.
 PR 09-FEB-1998: US-020956.
 PR 25-FEB-1997: US-806099.
 PR 01-AUG-1997: US-904804.
 PA (CORI-) CORIXA CORP.
 PI Dillon DC, Xu J;
 DR WPI: 98-609886/51.
 DR P-PSDB: W71871.
 PT Polypeptides comprising immunogenic portions of prostate proteins -
 PT used in a vaccine for the treatment of prostate cancer
 PS Claim 3: Page 104: 130pp; English.
 CC The present sequence is a new DNA which encodes an immunogenic portion
 CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
 CC can be used as a vaccine for the treatment of prostate cancer. The DNA
 CC was identified by analysis of a subtracted cDNA library obtained by
 CC subtracting a prostate tumour cDNA expression library with a normal
 CC tissue cDNA library.
 SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 30.3%; Score 338.8; DB 1; Length 1248;
 Best Local Similarity 93.9%; Pred. No. 3.4e-91;
 Matches 368; Conservative 0; Mismatches 12; Indels 12; Gaps 1;

QY 2 CGCAGCTGCGAGCCCTGGCAGGCGGCGACTGCTATGTAAGAAAGCAATGTTCTGCTGGGCG 61
 DB 35 CGCAGCTGCGAGCCCTGGCAGGCGGCGACTGCTATGTAAGAAAGCAATGTTCTGCTGGGCG 94
 QY 62 TCCGTGGTGCATCCGAGTGGGTGCTGTCAAGCCGACACTGTTTCCAGAA----- 111
 DB 95 TCCGTGGTGCATCCGAGTGGGTGCTGTCAAGCCGACACTGTTTCCAGAAAGTGAAGTGCAGA 154
 QY 111 -CTCTACACCATCGGCGGCTGCGTGCAGAGTCTTGAAGCGGACCAAGAGCGAGAGCC 169
 DB 155 GCTCTACACCATCGGCGGCTGCGTGCAGAGTCTTGAAGCGGACCAAGAGCGAGAGCC 214
 QY 170 AGATGTGAGAGCGCAGCTCTCGCTACGCGACCCAGATACAGACAGACCTTGTGCTGA 229
 DB 215 AGATGTGAGAGCGCAGCTCTCGCTACGCGACCCAGATACAGACAGACCTTGTGCTGA 274
 QY 230 ACGACCTCATGCTCATCAAGTTGAGCAATCCGTGCGAGTCTGACACCATCCGGAGCA 289
 DB 275 ACGACCTCATGCTCATCAAGTTGAGCAATCCGTGCGAGTCTGACACCATCCGGAGCA 334
 QY 290 TCAGCATGCTTTCAGAGTGCCTTACCGCGGGAACCTTGGCTGCTTGGTGGGCTC 349
 DB 335 TCAGCATGCTTTCAGAGTGCCTTACCGCGGGAACCTTGGCTGCTTGGTGGGCTC 394
 QY 350 TGGTGGCGAAGCATGCTGTGATTCATCCATCCAG 381
 DB 395 TGGTGGCGAAGCATGCTGTGATTCATCCATCCAG 426

RESULT 11
 ID X41114 standard; cDNA; 402 BP.
 AC X41114;
 DT 17-JUN-1999 (first entry)
 DE Human secreted protein 5' EST SEQ ID NO:58.
 KM Human; secreted protein; EST; expressed sequence tag; diagnosis;
 KM forensic; gene therapy; chromosome mapping; signal peptide;
 KM upstream regulatory sequence; cytokine activity; cell proliferation;
 KM differentiation; haematopoiesis regulation; tissue growth regulation;
 KM reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
 KM thrombolytic; anti-inflammatory; tumour inhibition; ds.
 OS Homo sapiens.
 PN WO9906548-A2.
 PD 11-FEB-1999.
 PF 31-JUL-1998: IB1222.
 PR 01-AUG-1997: US-905135.
 PA (GEST) GENSET.
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;

DR WPI: 99-153778/13.
DR P-PSDB: Y12281.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
PS Claim 1; Page 205; 824pp; English.
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 402 BP; 80 A; 125 C; 121 G; 75 T;

Query Match 30.2%; Score 338; DB 1; Length 402;

Best Local Similarity 96.8%; Pred. No. 3.4e-91; Mismatches 0; Indels 12; Gaps 1;
Matches 360; Conservative 0;

QY 2 CGCACTCCAGCCCTGCGAGGCGGCGATGTCATGTAAGAAACGATTTCTCTCGGCG 61
DB 23 CGCACTCCAGCCCTGCGAGGCGGCGATGTCATGTAAGAAACGATTTCTCTCGGCG 82
QY 62 TCCGTGTCATCCGAGTGGGTCTGTCAACCCGACACTGTTTCCAGAA-----111
DB 83 TCCGTGTCATCCGAGTGGGTCTGTCAACCCGACACTGTTTCCAGAAAGATKACA 142
QY 111 -CTCTACACCATGGGCTGGGCTGCGACAGCTTGAGCGCCGACCAAGACGAGGAGCC 169
DB 143 GCTCTTACACCATGGGCTGGGCTGCGACAGCTTGAGCGCCGACCAAGACGAGGAGCC 202
QY 170 AGATGGTGAGGCGCAGCTCTCCGATGAGGACCCAGAGTACAAAGACCTTGCCTGA 229
DB 203 AGATGGTGAGGCGCAGCTCTCCGATGAGGACCCAGAGTACAAAGACCTTGCCTGA 262
QY 230 AGACCTCATGCTCATCAAGTTGAGAGATCCGTCTCCGATGAGTACACCATCCGAGCA 289
DB 263 AGACCTCATGCTCATCAAGTTGAGAGATCCGTCTCCGATGAGTACACCATCCGAGCA 322
QY 290 TCAGATGTCCTTCGAGAGTCCCTACCGGGGGAATCTTTCCTCTTCTGCTGGGTC 349
DB 323 TCAGATGTCCTTCGAGAGTCCCTACCGGGGGAATCTTTCCTCTTCTGCTGGGTC 382
QY 350 TGCTGGCGAGC 361
DB 383 TGCTGGCGAGC 394

RESULT 12
V58646
ID V58646 standard; cDNA; 1459 BP.
AC V58646;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE6.
KW Prostate tumour specific gene; human; prostate cancer; detection;
OS Homo sapiens.
PN W09837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.

PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 114; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprising contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Query Match 23.0%; Score 257.4; DB 1; Length 1459;

Best Local Similarity 97.8%; Pred. No. 6.3e-67; Mismatches 6; Indels 0; Gaps 0;
Matches 261; Conservative 0;

QY 98 ACTGTTCCAGACTCTTACACATCCGAGCTGGGCTGCGACAGTCTTGGAGGCCGACCAAG 157
DB 25 AGTAGTGCANAGCTCTTACACATCCGAGCTGGGCTGCGACAGTCTTGGAGGCCGACCAAG 84
QY 158 AGCCAGGAGGCCAGATGGTGGAGGCCAGCTCTCCGTAAGGACCCAGAGTACAAACAGAC 217
DB 85 AGCCAGGAGGCCAGATGGTGGAGGCCAGCTCTCCGTAAGGACCCAGAGTACAAACAGAC 144
QY 218 CCTGTGCTGCTAAGACCTCATGCTCATCAAGTTGAGAGCAATCCGTCCGAGTCTACA 277
DB 145 CTTGTGCTGCTAAGACCTCATGCTCATCAAGTTGAGAGCAATCCGTCCGAGTCTACA 204
QY 278 CCATCCGAGGAGCATGAGTCTTCCGAGTCCCTTACCGGGGGAATCTTGCCTGTTT 337
DB 205 CCATCCGAGGAGCATGAGTCTTCCGAGTCCCTTACCGGGGGAATCTTGCCTGTTT 264
QY 338 CTGGCTGGGGTCTGCTGGCGAAGCATG 364
DB 265 CTGGCTGGGGTCTGCTGGCGAAGCATG 291

RESULT 13

V61251
ID V61251 standard; cDNA; 1459 BP.

AC V61251;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE6.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN W09837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 106; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:34:10 ; Search time 161.06 Seconds
(without alignments)
636.943 Million cell updates/sec

Title: US-09-030-606-177

Perfect score: 1119

Sequence: 1 GCCGACCTCGCAGCCCTGCA.....ATGTTAAAAA 1119

Scoring table: IDENTITY_NUC

Searched: 176461 seqs, 45838279 residues

Database :

1: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/1na/5C.COMB.seq:*
4: /cgn2_6/ptodata/2/1na/5D.COMB.seq:*
5: /cgn2_6/ptodata/2/1na/PTUS9.COMB.seq:*
6: /cgn2_6/ptodata/2/1na/Backfillseq1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	350.4	31.3	871	3	US-08-744-026-2
2	81	7.2	986	4	US-08-557-146-1
3	81	7.2	1089	5	PCT-US96-04294-1
4	81	7.2	1089	5	PCT-US96-04294-3
5	76.4	6.8	832	5	PCT-US95-06157-5
6	72.8	6.5	992	2	US-08-358-782D-13
7	72.6	6.5	1462	2	US-08-358-782D-14
8	72	6.4	200	3	US-08-931-981A-1
9	71.6	6.4	766	5	PCT-US95-06157-9
10	68	6.1	1729	4	US-08-844-024-1
11	63.6	5.7	760	5	PCT-US95-06157-7
12	62.4	5.6	833	4	US-08-790-137-2
13	58.4	5.2	2259	4	US-08-845-998-3
14	56.8	5.1	2218	4	US-08-845-998-5
15	50.8	4.5	957	3	US-08-684-862-11
16	49.6	4.4	1454	3	US-08-467-155A-2
17	49.6	4.4	1454	4	US-08-628-198-2
18	49.6	4.4	1454	5	PCT-US96-07343-2
19	46	4.1	840	3	US-08-684-862-12
20	45.6	4.1	732	3	US-08-361-395-2
21	45	4.0	111	3	US-08-454-720A-42
22	40.8	3.6	925	1	US-07-990-301A-3
23	40.4	3.6	7218	2	US-08-232-463-14
24	40.2	3.6	734	2	US-08-650-129-1
25	40.2	3.6	821	3	US-08-650-129-2
26	40.2	3.6	866	3	US-08-650-129-3
27	38.4	3.4	1096	3	US-08-684-862-8
28	36.4	3.3	1554	3	US-08-469-486-1
29	36	3.2	970	1	US-08-148-910-3
30	36	3.2	2033	1	US-08-148-910-14
31	36	3.2	970	2	US-08-448-937A-3
32	36	3.2	2033	2	US-08-448-937A-14
33	35.8	3.2	959	3	US-08-568-031-1
34	35.8	3.2	959	4	US-08-966-319-1
35	35.4	3.2	2743	1	US-08-396-479B-3
36	35.4	3.2	2743	2	US-08-818-823-3
37	35.4	3.2	2749	4	US-08-124-981A-1

ALIGNMENTS

38	35.4	3.2	2853	5	PCT-US94-07297-36	Sequence 36, Appl
39	35	3.1	807	2	US-08-270-584A-1	Sequence 1, Appl
40	35	3.1	807	4	US-08-765-192-1	Sequence 1, Appl
41	34.8	3.1	867	1	US-07-990-301A-1	Sequence 1, Appl
42	34.2	3.1	1333	3	US-08-684-862-9	Sequence 1, Appl
43	33.8	3.0	681	1	US-07-929-198-1	Sequence 1, Appl
44	33.8	3.0	681	1	US-07-929-198-3	Sequence 3, Appl
45	33.8	3.0	681	1	US-07-929-198-5	Sequence 5, Appl

RESULT 1
US-08-744-026-2
Sequence 2, Application US/08744026
Patent No. 5786148
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
TITLE OF INVENTION: KALIKREIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/744,026
FILING DATE: Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0154 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 871 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY:
CLONE: Consensus
US-08-744-026-2

Query Match 31.3%; Score 350.4; DB 3; Length 871;
Best Local Similarity 95.0%; Pred. No. 5.6e-103;
Matches 360; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 3 GCAATCCAGCCCGGAGGCGGACATGTCATGGAAGCAATGTTGCTGGCGGT 62
DB 129 GCAATCCAGCCCGGAGGCGGACATGTCATGGAAGCAATGTTGCTGGCGGT 188
QY 63 CTTGTCATCCGAGGAGTGGTGTCTGTCAGCCGACACTGTTCCAGAACTCTACACCAT 122
DB 189 CTTGTCATCCGAGGAGTGGTGTCTGTCAGCCGACACTGTTCCAGAACTCTACACCAT 248

Oy	123	CGGGCTGGGCGCTGCACAACTCTTTGAGCGGACCAAGAGCCAGAGGAGCCAGATGGTGGAGGC	182
Db	249	CGGGCTGGGCGCTGCACAACTCTTTGAGCGGACCAAGAGCCAGAGGAGCCAGATGGTGGAGGC	308
Oy	183	CAGCCTCTCCGTAGGGACCCAGAGTACACAGACCCCTTGCTGTCTAAGACCTCATGCT	242
Db	309	CAGCCTCTCCGTAGGGACCCAGAGTACACAGACCCCTTGCTGTCTAAGACCTCATGNT	368
Oy	243	CATCAAGTTGGACGATCCGTGCCGAGTCTGACACCATCCGGAGCATACAGANTGGTTC	302
Db	359	CATCAAGTTGGACGATCCGTGCCGAGTCTGACAAACATCCGGAGNATCAGCATTTGTTTC	428
Oy	303	GCAAGTGCCCTACCGGGGGGAACTCTTGCTGTTCTTGCGTGGGGTGTGCTGGCGAACA	362
Db	429	GCAGGCGCCCTACCGGGGGGAACTTTTGCTGTCTTGCGTGGGGGTGTGCTGGCGAACG	488
Oy	363	TGCTGTGATTCGATCCGACG	381
Db	489	CAGAAATGCTTACCGTGGTG	507

```

1      RESULT      2
2      US-08-557-146-1
3      ; Sequence 1, Application US/08557146
4      ; Patent No. 5834290
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Egeldur, Torbjorn
7      ; APPLICANT: Hansson, Lennart
8      ; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
9      ; TITLE OF INVENTION: Enzyme (SCCE)
10     ; NUMBER OF SEQUENCES: 17
11     ; CORRESPONDENCE ADDRESS:
12     ; ADDRESSEE: White & Case, Patent Department
13     ; STREET: 1155 Avenue of the Americas
14     ; CITY: New York
15     ; STATE: New York
16     ; COUNTRY: U.S.A.
17     ; ZIP: 10036-2787
18     ; COMPUTER READABLE FORM:
19     ; MEDIUM TYPE: Floppy disk
20     ; COMPUTER: IBM PC compatible
21     ; OPERATING SYSTEM: PC-DOS/MS-DOS
22     ; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
23     ; CURRENT APPLICATION DATA:
24     ; APPLICATION NUMBER: US/08/557,146
25     ; FILING DATE: 14-DEC-1995
26     ; CLASSIFICATION: 424
27     ; ATTORNEY/AGENT INFORMATION:
28     ; NAME: Steiner, Richard J.
29     ; REGISTRATION NUMBER: 35,372
30     ; REFERENCE/DOCKET NUMBER: 1103326-181
31     ; TELECOMMUNICATION INFORMATION:
32     ; TELEPHONE: (212) 354-8113
33     ; TELEFAX: (212) 354-8113
34     ; INFORMATION FOR SEQ ID NO: 1:
35     ; SEQUENCE CHARACTERISTICS:
36     ; LENGTH: 986 base pairs
37     ; TYPE: nucleic acid
38     ; STRANDEDNESS: single
39     ; TOPOLOGY: linear
40     ; MOLECULE TYPE: cDNA
41     ; HYPOTHETICAL: NO
42     ; ANTI-SENSE: NO
43     ; ORIGINAL SOURCE:
44     ; ORGANISM: Homo sapiens
45     ; FEATURE:
46     ; NAME/KEY: CDS
47     ; LOCATION: 25..786
48     ; FEATURE:
49     ; NAME/KEY: sig_peptide
50     ; LOCATION: 25..90
51     ; FEATURE:
52     ; NAME/KEY: mat_peptide
53     ;
54     ;
55     ;
56     ;
57     ;
58     ;
59     ;
60     ;
61     ;
62     ;
63     ;
64     ;
65     ;
66     ;
67     ;
68     ;
69     ;
70     ;
71     ;
72     ;
73     ;
74     ;
75     ;
76     ;
77     ;
78     ;
79     ;
80     ;
81     ;
82     ;
83     ;
84     ;
85     ;
86     ;
87     ;
88     ;
89     ;
90     ;
91     ;
92     ;
93     ;
94     ;
95     ;
96     ;
97     ;
98     ;
99     ;
100    ;
101    ;
102    ;
103    ;
104    ;
105    ;
106    ;
107    ;
108    ;
109    ;
110    ;
111    ;
112    ;
113    ;
114    ;
115    ;
116    ;
117    ;
118    ;
119    ;
120    ;
121    ;
122    ;
123    ;
124    ;
125    ;
126    ;
127    ;
128    ;
129    ;
130    ;
131    ;
132    ;
133    ;
134    ;
135    ;
136    ;
137    ;
138    ;
139    ;
140    ;
141    ;
142    ;
143    ;
144    ;
145    ;
146    ;
147    ;
148    ;
149    ;
150    ;
151    ;
152    ;
153    ;
154    ;
155    ;
156    ;
157    ;
158    ;
159    ;
160    ;
161    ;
162    ;
163    ;
164    ;
165    ;
166    ;
167    ;
168    ;
169    ;
170    ;
171    ;
172    ;
173    ;
174    ;
175    ;
176    ;
177    ;
178    ;
179    ;
180    ;
181    ;
182    ;
183    ;
184    ;
185    ;
186    ;
187    ;
188    ;
189    ;
190    ;
191    ;
192    ;
193    ;
194    ;
195    ;
196    ;
197    ;
198    ;
199    ;
200    ;
201    ;
202    ;
203    ;
204    ;
205    ;
206    ;
207    ;
208    ;
209    ;
210    ;
211    ;
212    ;
213    ;
214    ;
215    ;
216    ;
217    ;
218    ;
219    ;
220    ;
221    ;
222    ;
223    ;
224    ;
225    ;
226    ;
227    ;
228    ;
229    ;
230    ;
231    ;
232    ;
233    ;
234    ;
235    ;
236    ;
237    ;
238    ;
239    ;
240    ;
241    ;
242    ;
243    ;
244    ;
245    ;
246    ;
247    ;
248    ;
249    ;
250    ;
251    ;
252    ;
253    ;
254    ;
255    ;
256    ;
257    ;
258    ;
259    ;
260    ;
261    ;
262    ;
263    ;
264    ;
265    ;
266    ;
267    ;
268    ;
269    ;
270    ;
271    ;
272    ;
273    ;
274    ;
275    ;
276    ;
277    ;
278    ;
279    ;
280    ;
281    ;
282    ;
283    ;
284    ;
285    ;
286    ;
287    ;
288    ;
289    ;
290    ;
291    ;
292    ;
293    ;
294    ;
295    ;
296    ;
297    ;
298    ;
299    ;
300    ;
301    ;
302    ;
303    ;
304    ;
305    ;
306    ;
307    ;
308    ;
309    ;
310    ;
311    ;
312    ;
313    ;
314    ;
315    ;
316    ;
317    ;
318    ;
319    ;
320    ;
321    ;
322    ;
323    ;
324    ;
325    ;
326    ;
327    ;
328    ;
329    ;
330    ;
331    ;
332    ;
333    ;
334    ;
335    ;
336    ;
337    ;
338    ;
339    ;
340    ;
341    ;
342    ;
343    ;
344    ;
345    ;
346    ;
347    ;
348    ;
349    ;
350    ;
351    ;
352    ;
353    ;
354    ;
355    ;
356    ;
357    ;
358    ;
359    ;
360    ;
361    ;
362    ;
363    ;
364    ;
365    ;
366    ;
367    ;
368    ;
369    ;
370    ;
371    ;
372    ;
373    ;
374    ;
375    ;
376    ;
377    ;
378    ;
379    ;
380    ;
381    ;
382    ;
383    ;
384    ;
385    ;
386    ;
387    ;
388    ;
389    ;
390    ;
391    ;
392    ;
393    ;
394    ;
395    ;
396    ;
397    ;
398    ;
399    ;
400    ;
401    ;
402    ;
403    ;
404    ;
405    ;
406    ;
407    ;
408    ;
409    ;
410    ;
411    ;
412    ;
413    ;
414    ;
415    ;
416    ;
417    ;
418    ;
419    ;
420    ;
421    ;
422    ;
423    ;
424    ;
425    ;
426    ;
427    ;
428    ;
429    ;
430    ;
431    ;
432    ;
433    ;
434    ;
435    ;
436    ;
437    ;
438    ;
439    ;
440    ;
441    ;
442    ;
443    ;
444    ;
445    ;
446    ;
447    ;
448    ;
449    ;
450    ;
451    ;
452    ;
453    ;
454    ;
455    ;
456    ;
457    ;
458    ;
459    ;
460    ;
461    ;
462    ;
463    ;
464    ;
465    ;
466    ;
467    ;
468    ;
469    ;
470    ;
471    ;
472    ;
473    ;
474    ;
475    ;
476    ;
477    ;
478    ;
479    ;
480    ;
481    ;
482    ;
483    ;
484    ;
485    ;
486    ;
487    ;
488    ;
489    ;
490    ;
491    ;
492    ;
493    ;
494    ;
495    ;
496    ;
497    ;
498    ;
499    ;
500    ;
501    ;
502    ;
503    ;
504    ;
505    ;
506    ;
507    ;
508    ;
509    ;
510    ;
511    ;
512    ;
513    ;
514    ;
515    ;
516    ;
517    ;
518    ;
519    ;
520    ;
521    ;
522    ;
523    ;
524    ;
525    ;
526    ;
527    ;
528    ;
529    ;
530    ;
531    ;
532    ;
533    ;
534    ;
535    ;
536    ;
537    ;
538    ;
539    ;
540    ;
541    ;
542    ;
543    ;
544    ;
545    ;
546    ;
547    ;
548    ;
549    ;
550    ;
551    ;
552    ;
553    ;
554    ;
555    ;
556    ;
557    ;
558    ;
559    ;
560    ;
561    ;
562    ;
563    ;
564    ;
565    ;
566    ;
567    ;
568    ;
569    ;
570    ;
571    ;
572    ;
573    ;
574    ;
575    ;
576    ;
577    ;
578    ;
579    ;
580    ;
581    ;
582    ;
583    ;
584    ;
585    ;
586    ;
587    ;
588    ;
589    ;
590    ;
591    ;
592    ;
593    ;
594    ;
595    ;
596    ;
597    ;
598    ;
599    ;
600    ;
601    ;
602    ;
603    ;
604    ;
605    ;
606    ;
607    ;
608    ;
609    ;
610    ;
611    ;
612    ;
613    ;
614    ;
615    ;
616    ;
617    ;
618    ;
619    ;
620    ;
621    ;
622    ;
623    ;
624    ;
625    ;
626   
```

LOCATION: 112..783
US-08-557-146-1

Query Match 7.2%; Score 81; DB 4; Length 986;
Best Local Similarity -53.7%; Pred. No. 1.8e-16;
Matches 196; Conservative 0; Mismatches 160; Indels 9; Gaps 1;

Oy	6	CTCCAGGCCCCGCGAGGCGGCGACTGGTCAATGGAAAACGAATTTGTCGTCGGGGCGTCCT	65
Db	141	CTCCACCCCATGTGCGAGGTGGCCCTTGCTCAATGGCAATCACTCCACTCGCGAGGCGCTCT	200
Oy	66	GGTGCATCCGCAGTGGGTGCTGTGCACCGGCACACTGTTTCCAGAACTCCTACACCATCGG	125
Db	201	GGTCAATGAGCGCTGGGTGGTCACTGGCGGCCACTGCAAGATGAATGATGACACCTGCA	260
Oy	126	GCTGGGCGCTGCACAGTTTGAAGGCCGACCAAGAACCGAGGAGCCAGATGGTGGAGGCCAG	185
Db	261	CCTGGCGAGTATACGCTGGGCGCACAGAGAG-----CTCAGAGGATCAAGGCGCTC	311
Oy	186	CCTCTCGGTAGGGAGACCCAGAGATCAACACAAACCCCTTGCTGCTAAGAGACTATGCTCAT	245
Db	312	GAAGTCAATTCGGCCACCCCGGCTACTCCACAGAACCCATGTTAATGACTCATGGTCTGT	371
Oy	246	CAAGTTGGACGAATCCGTGTCGAGTGTGACACACATCCGAGGATCAGATTCCTTCGCA	305
Db	372	GAACTCAATATGCGACAGGCCAAGCTGTATCTCATGTGTAAAGAAAGTCAAGCTGCCCTCCGG	431
Oy	306	GTGCGCTTACCGCGGGGGAATCTTGCCCTGTTGCTTGCGTGGGGGTCTGCTGGCGAAGATGC	365
Db	432	CTGGGAAGCCCTGGAAACACACTGTACTGTCCTCGGCTGGGGCACTAACACGAGCCAG	491
Oy	366	TGTGA 370	
Db	492	TGTGA 496	

RESULT 3
 PCT-US96-04294-1/c
 Sequence 1, Application PC/TUS9604294
 GENERAL INFORMATION:
 APPLICANT: Dixon, Eric P.
 APPLICANT: Johnstone, Edward M.
 APPLICANT: Little, Shella P.
 TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEINASE AND
 TITLE OF INVENTION: RELATED NUCLEIC ACIDS
 NUMBER OF SEQUENCES: 3
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Eli Lilly and Company
 STREET: Lilly Corporate Center
 CITY: Indianapolis
 STATE: Indiana
 COUNTRY: United States of America
 ZIP: 46285
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US96/04294
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/416,257
 FILING DATE: 04-APR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Bialock, Donna K.
 REGISTRATION NUMBER: 38,082
 REFERENCE/DOCKET NUMBER: X92339
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 317-277-1090
 TELEFAX: 317-276-3861

Query Match 6.5%; Score 72.8; DB 2; Length 992;
Best Local Similarity 59.9%; Pred. No. 7.7e-14;
Matches 139; Conservative 0; Mismatches 89; Indels 4; Gaps 1

QY	303	GCAATGCCCCACCCGGGGGAACCTTTCCCTCGTTTCGCGTGGGGGTTCGCTGGCGAACA	362
Db	31	GCTGTGNNACCAATGTGTGGGTCCCGGGTGTCTTCTCCACCTGTCCCTGACGTGGATTGG	90
QY	363	TGCTGTGATTGCATCCAGTCCACAGACTGTGGAGAGCGTGGAGTGTGAGAAAGCTTTCCA	422
Db	91	TGCTGCACCCCTCATCCCTGCTCGTCCGATTGTGGAGCGCTGGGAGTGGCAAGCATTTCCA	150
QY	423	ACCCCTGGCAGG-----GTTGTACCATTTTCGGCACTTCCAGTGCAGAAGACGTCTGCTGCA	478
Db	151	ACCCCTGGGAGGTGCTTGTAGCCCTTCGTGGAGGGCAGATGTCCGGCCGTTCTTGTGTGCA	210
QY	479	TTCATCATGGGTGCTCACTACTGTCACAGCATACCCGGAAACATCTGTGAT	530
Db	211	CCCCAGTGGTCTCTCAAGCTGCCACACTGATATAGCAACAAAGCCTGTATC	262

RESULT 7
 : Sequence 14, Application US/08358782D
 : Patent No. 5674682
 : GENERAL INFORMATION:
 : APPLICANT: Croce, Carlo
 : APPLICANT: Gomella, Leonard
 : APPLICANT: Mulholland, S. Grant
 : APPLICANT: Moreno, Jose
 : APPLICANT: Fischer, Rainer
 : TITLE OF INVENTION: Methods of Detecting Micrometastasis of Prostate
 : NUMBER OF SEQUENCES: 14
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
 : STREET: One Liberty Place 46th. Floor
 : CITY: Philadelphia
 : STATE: PA
 : ZIP: 19103
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/358,782D
 : FILING DATE: 15-DEC-1994
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Beardell, Lori Y.
 : REGISTRATION NUMBER: 34,293
 : REFERENCE/DOCKET NUMBER: TJU-1327
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 215-568-3100
 : TELEFAX: 215-568-3439

US-08-931-981A-1

Matches	109;	conservative	0;	Mismatches	39;	Indels	4;	Gaps	1,
---------	------	--------------	----	------------	-----	--------	----	------	----

3/5 CATTCCAGTCCAGCTGTGGAGGCTGGAGTGTGAGGAAGCTTCCCAACCTGCGAGG-43

US-08-684-862-11

Query Match 4.58; Score 50.8; DB 3; Length 957;
Best Local Similarity 48.08; Pred. No. 8.7e-07;
Matches 145; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

```
QY 46 TTGTTCTGCTCGGCGCTCTGTCATCCGCGTGGTGTGTCAGCCGACACTGTTTC 105
   || || || || || || || || || || || || || || || || || || || ||
Db 279 TTTGTCTGCGCTGGGACTTTGATCCACCAGAAATGGGTGTCACCTGTCACACTGCGAG 338
   || || || || || || || || || || || || || || || || || || || ||
QY 106 CAGAACTCTTACCACTGCGGCTGGGCTGCGACAGCTTTGAGGCCGACCAAGAGCCAGGG 165
   || || || || || || || || || || || || || || || || || || || ||
Db 339 AGTACGGATCTCAAGATGAAGTTGTATGCATAGCAAAAAGGTACAAAATAGAGATGAG 398
   || || || || || || || || || || || || || || || || || || || ||
QY 166 AGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGACCCAGAGTACAAAGAGACCCTTGCTC 225
   || || || || || || || || || || || || || || || || || || || ||
Db 399 CAGACAAAGAAACGCAAGAAAGTTTCTATTGTCCCAATTAAGAAAACGATGAAGTACTG 458
   || || || || || || || || || || || || || || || || || || || ||
QY 226 GCTAACGACCTCATGCTCATCAATGAGTGGAGCAATCCGTGTCGAGTCTGACACCATCCGG 285
   || || || || || || || || || || || || || || || || || || || ||
Db 459 GACAAAGACATTAATGATGATCAAGCTGACCACTCTGTAGCAATAGTAAACACATCGCG 518
   || || || || || || || || || || || || || || || || || || || ||
QY 286 AGCATCAGCATGCTTGGCAGTGGCCTTACCGGGGGAACCTTGCCTGCTTTCGTGGCTGG 345
   || || || || || || || || || || || || || || || || || || || ||
Db 519 CCTCTCAGCTTGCTTCCAGCCCTCCAGGTGTGGGCTCATTTTGCATATTATGGGATGG 578
   || || || || || || || || || || || || || || || || || || || ||
QY 346 GG 347
   || ||
Db 579 GG 580
```

Search completed: September 28, 1999, 11:34:11
Job time: 2131 sec

This Page Blank (uspto)

OM of: US-09-030-606-177 to: SwissProt_37:* out_format : pfs
Date: Sep 25, 1999 4:46 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:

-MODEL=frame+2p.model -DEV=xlp
-O=/gen2.1/USPRO_pool/US09030606/runat_24091999_171618_29683/app-query.fasta.1
-DB=SwissProt_37 -OFM=fastan -SUFFIX=isp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-GAPEXT=0.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-TRANS=human40.cdt -LIST=45 -DOCLIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPD=6 -ICPD=3 -WAIT -THREADS=1

Search information block:

Query: US-09-030-606-177
Query length: 1119
Database: SwissProt_37:*
Database sequences: 77977
Database length: 28268293
Search time (sec): 155.430000

score_list:

Sequence	Strd Orig	ZScore	Escore	len	Documentation
SwissProt_37:KLK1_MACFA	275.00	435.48	4.2e-17	257	P00776 macaca fascicularis
SwissProt_37:SCCE_HUMAN	274.50	434.82	4.6e-17	253	P49652 homo sapiens (human)
SwissProt_37:KLK1_HUMAN	272.50	431.25	7.0e-17	262	P06870 homo sapiens (human)
SwissProt_37:KLK1_PAPHA	270.50	428.15	1.1e-16	258	P02873 papio hamadryas (hama
SwissProt_37:PROS_HUMAN	270.00	427.23	1.2e-16	261	P07288 homo sapiens (human)
SwissProt_37:KLK2_HUMAN	266.50	421.56	2.4e-16	261	P20151 homo sapiens (human)
SwissProt_37:PROS_MACMU	262.00	414.27	6.2e-16	261	P33619 macaca mulatta (rhesu
SwissProt_37:KLK1_MOUSE	255.00	411.10	9.4e-16	259	P15948 mus musculus (mouse)
SwissProt_37:KLK1_MOUSE	255.00	402.92	2.7e-15	261	P00756 mus musculus (mouse)
SwissProt_37:KLK1_MOUSE	253.50	401.12	3.6e-15	244	P36375 rattus norvegicus (rat)
SwissProt_37:KLK1_MOUSE	250.50	395.63	6.8e-15	261	P15946 mus musculus (mouse)
SwissProt_37:TRRY_CHICK	249.00	393.68	9.2e-15	248	P09029 gallus gallus (chick)
SwissProt_37:TRRY_CANFA	248.00	392.13	1.1e-14	246	P06871 canis familiaris (dog
SwissProt_37:KLK2_RAT	246.50	389.92	1.6e-14	259	P00759 rattus norvegicus (rat)
SwissProt_37:KLK1_RAT	245.50	387.53	1.9e-14	261	P36373 rattus norvegicus (rat)
SwissProt_37:TRRY_RAT	245.00	387.27	2.1e-14	246	P32821 rattus norvegicus (rat)
SwissProt_37:TRRY_BOVIN	245.00	387.27	2.1e-14	246	P32822 rattus norvegicus (rat)
SwissProt_37:TRRY_HUMAN	244.00	385.76	2.6e-14	243	P00760 bos taurus (bovine)
SwissProt_37:KLK1_RAT	244.00	385.10	2.6e-14	261	P00758 rattus norvegicus (rat)
SwissProt_37:KLK1_MOUSE	244.00	385.10	2.6e-14	261	P15949 mus musculus (mouse)
SwissProt_37:KLK1_MOUSE	243.50	384.47	2.9e-14	256	P00757 mus musculus (mouse)
SwissProt_37:KLK1_MOUSE	243.50	384.29	2.9e-14	261	P36374 rattus norvegicus (rat)
SwissProt_37:TRRY_SODAC	243.00	384.70	3.1e-14	229	P00764 squalus acanthias (sh
SwissProt_37:TRRY_HUMAN	243.00	383.99	3.2e-14	247	P08426 rattus norvegicus (rat)
SwissProt_37:TRRY_HUMAN	242.50	381.25	3.7e-14	304	P35030 homo sapiens (human)
SwissProt_37:KLK1_MOUSE	242.00	381.86	4.0e-14	261	P36369 mus musculus (mouse)
SwissProt_37:TRRY_RAT	241.00	380.79	4.8e-14	246	P00762 rattus norvegicus (rat)
SwissProt_37:KLK1_MOUSE	241.00	380.24	4.9e-14	261	P36368 mus musculus (mouse)
SwissProt_37:KLK1_MOUSE	240.50	379.50	5.4e-14	259	P36376 rattus norvegicus (rat)
SwissProt_37:TRRY_HUMAN	240.00	379.13	6.0e-14	247	P15951 homo sapiens (human)
SwissProt_37:TRRY_RAT	240.00	379.13	6.0e-14	247	P12788 rattus norvegicus (rat)
SwissProt_37:TRRY_CHICK	239.00	377.47	7.3e-14	248	P09627 gallus gallus (chick)
SwissProt_37:KLK1_CANPO	239.00	377.47	7.3e-14	248	P09628 gallus gallus (chick)
SwissProt_37:TRRY_HUMAN	237.50	375.39	1.0e-13	239	P12323 canis familiaris (dog)
SwissProt_37:TRRY_MOUSE	237.00	374.31	1.1e-13	246	P00746 mus musculus (mouse)
SwissProt_37:TRRY_HUMAN	237.00	374.31	1.1e-13	247	P07447 mus musculus (mouse)
SwissProt_37:TRRY_HUMAN	237.00	373.27	1.1e-13	260	P09582 canis familiaris (dog)
SwissProt_37:TRRY_XENLA	236.00	372.76	1.4e-13	244	P70059 xenopus laevis (frog)
SwissProt_37:KLK1_MOUSE	236.00	372.13	1.4e-13	261	P15945 mus musculus (mouse)
SwissProt_37:TRRY_PIG	235.00	371.65	1.7e-13	231	P00761 sus scrofa (pig)
SwissProt_37:KLK1_MOUSE	234.00	368.89	2.1e-13	261	P15947 mus musculus (mouse)

SwissProt_37:KLK1_MOUSE + 234.00 368.89 2.1e-13 261 P00755 mus musculus (mous
SwissProt_37:TRRY_SALSA + 233.00 366.79 3.1e-13 231 P35032 salmo salar (atlan
SwissProt_37:KLK1_MOUSE + 231.00 364.03 3.9e-13 261 P07628 mus musculus (mous

seq_name: SwissProt_37:KLK1_MACFA

seq_documentation_block:

ID	KLK1_MACFA	STANDARD	PRT	257 AA
AC	007276:			
DT	01-OCT-1994 (REL. 30, CREATED)			
DT	01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)			
DT	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)			
DE	GLANDULAR KALLIKREIN 1 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)			
DE	(KIDNEY/PANCREAS/SALIVARY GLAND KALLIKREIN)			
GN	KLK1			
OS	MACACA FASCICULARIS (CRAB EATING MACAQUE) (CYNOMOLGUS MONKEY)			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;			
OC	PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; MACACA.			
NC	[1]			
RP	SEQUENCE FROM N.A.			
RA	LIN F.K., LIN C.H., CHOU C., CHEN K., LU H.S., BACHELLER B.,			
RA	HERRERA C., JONES T., CHAO J., CHAO L.;			
RT	"Molecular cloning and sequence analysis of the monkey and human			
RT	tissue kallikrein genes."			
RL	BIOCHIM. BIOPHYS. ACTA 1173:325-328(1993).			
CC	- FUNCTION: GLANDULAR KALLIKREIN CLEAVE MET-LYS AND ARG-SER BONDS			
CC	IN KININOGEN TO RELEASE LYS-BRADYKININ.			
CC	- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-1-XAA BONDS IN			
CC	SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE			
CC	KALLIDIN (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF			
CC	MET-1-XAA OR LEU-1-XAA.			
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1: ALSO KNOWN AS THE			
CC	TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: L10039; G293141; -			
DR	PIR: S33772; S33772			
DR	PROSITE: P500134; TRYPsin_HIS; 1.			
DR	PROSITE: P500135; TRYPsin_SER; 1.			
DR	PFAM: PF00089; trypsin; 1.			
DR	HSSP: P00752; IHI.			
KW	HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN; PANCREAS;			
KW	ZYMOGEN; SIGNAL.			
FT	SIGNAL	1	18	BY SIMILARITY.
FT	PROPEP	19	24	ACTIVATION PEPTIDE (PROBABLE).
FT	CHAIN	25	257	GLANDULAR KALLIKREIN 1.
FT	ACT SITE	62	62	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT SITE	116	116	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT SITE	209	209	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	CARBOHYD	90	90	POTENTIAL.
FT	CARBOHYD	99	99	POTENTIAL.
FT	CARBOHYD	99	99	POTENTIAL.
FT	CARBOHYD	101	101	POTENTIAL.
FT	CARBOHYD	105	105	POTENTIAL.
FT	CARBOHYD	160	160	POTENTIAL.
FT	CARBOHYD	162	162	POTENTIAL.
FT	DISULFID	31	169	BY SIMILARITY.
FT	DISULFID	47	63	BY SIMILARITY.
FT	DISULFID	148	215	BY SIMILARITY.
FT	DISULFID	180	194	BY SIMILARITY.
FT	DISULFID	205	230	BY SIMILARITY.
FT	SEQUENCE	257 AA;	28237 MW;	82A709DC CRC32;

alignment_scores: 275.00

Length: 158

Ratio: 2.865 Gaps: 3
Percent Similarity: 60.759 Percent Identity: 37.342

alignment block:
US-09-030-606-177 x KLK1_MAFRA

Align seg 1/1 to: KLK1_MAFRA from: 1 to: 257

```

7 TCGCAGCCCTGGAGCGGCGGAGTGTGATGGAAGAAACGAAATGTGTCTGTC 56
|||||
32 SerGlnProThrPrlGlnAlaLeuTyrHisPheSerThrPheGlnCysG1 48
|||||
57 GGGCGTCTGGTGTGATCCGAGTGGTGGTGTGACCGGACACATGTTTCC 106
|||||
48 YglYleuValHisProGlnThrValLeuThrAlaAlaHisCysG1 65
|||||
107 AGAACCTCTACACATGGGGCTGGCGCTGACAGTCTTGAGCGGACCAA 156
|||||
65 etaspsnryrGlnLeuThrPleuGlnYargHisAsnLeu...PheAsp 80
|||||
157 GAGCCAGGAGGAGCAGATGTGTGAGCGGCGCTCTCCGTACGGCAGCAGA 206
|||||
81 GlaspThrAlaGlnPheValHisValSerGluSerPheProHisProG1 97
|||||
207 GTACACAGACCCCTGTGCTGCTAAC..... 231
|||||
97 yPheAsnMetSerLeuLeuLysAsnHisThrArgGlnAlaAspArg 114
|||||
232 .....GACCTCATGCTCATCAAGTGTGAGGAATCCGCTCCAGTCTGAC 276
|||||
114 etHisAspLeuMetLeuLeuArgLeuThrGlnProAlaGlnLeuHis 130
|||||
277 ACCATCGGAGCATCATCAATGTGTGCTGCTGCTACCGGAGGAGTCTG 326
|||||
131 AlaValGlnValValGlnLeuProThrGlnGlnProGlnValGlySer 147
|||||
327 TTGGCTTCGTTTGGCTGGGCT..... 348
|||||
147 rcYsLeuAlaSerGlyTyrGlySerIleGluProGluAsnPheSerPhe 164
|||||
349 .....CTGGTGGGAGAGCATGCTG 369
|||||
164 roAspAspLeuGlnCysValAspLeuGlnLeuProAsnAspGluCys 180
|||||
370 ATTGCATCCAGTCCGAGACTGTG 393
|||||
181 AlalysAlaHisThrGlnLysVal 188
|||||
seq_name: SwissProt_37:SCCE_HUMAN
seq_documentation_block:
AC P49862;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE STRATUM CORNEUM CHYMOTRYPTIC ENZYME PRECURSOR (EC 3.4.21.-) (SCCE).
OS PRS6 OR SCCE.
GN HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.
RC TISSUE=SKIN;
RX MEDLINE: 94308225.
RA HANSON L., STROEMOVIST M., BACKMAN A., WALLBRANDT P., CARLSTEIN A.,
RA EGERLUD T.,
"Cloning, expression, and characterization of stratum corneum
RT chymotryptic enzyme. A skin-specific human serine proteinase."
RL J. BIOL. CHEM. 269:19420-19426(1994).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE: 95314630.

```

```

RA SKYT A., STROEMOVIST M., EGERLUD T.;
RT "Primary substrate specificity of recombinant human stratum corneum
RT chymotryptic enzyme."
RL BIOCHEM. BIOPHYS. RES. COMMUN. 211:586-589(1995).
CC -1- FUNCTION: MAY CATALYZE THE DEGRADATION OF INTERCELLULAR COHESIVE
CC STRUCTURES IN THE CORNIFIED LAYER OF THE SKIN IN THE CONTINUOUS
CC SHEDDING OF CELLS FROM THE SKIN SURFACE. SPECIFIC FOR AMINO ACID
CC RESIDUES WITH AROMATIC SIDE CHAINS IN THE P1 POSITION. SCCE
CC CLEAVES INSULIN B CHAIN AT 6-LEU-1-CYS-7, 16-TYR-1-LEU-17, 25-PHE-
CC -1-TYR-26, AND 26-TYR-1-THR-27. COULD PLAY A ROLE IN THE ACTIVATION
CC OF PRECURSORS TO INFLAMMATORY CYTOKINES.
CC -1- TISSUE SPECIFICITY: IT IS ABUNDANTLY EXPRESSED IN THE SKIN AND IS
CC EXPRESSED BY KERATINOCYTES IN THE EPIDERMIS. VERY LOW LEVELS ARE
CC ALSO SEEN IN THE BRAIN AND KIDNEY.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L33404; G532504;
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PRM: PF00089; trypsin; 1.
DR HSSP: P00763; IDPO.
KW HYDROLASE; SERINE PROTEASE; ZMOGEN; GLYCOPROTEIN; SIGNAL.
FT SIGNAL 23
FT PROPEP 1
FT CHAIN 30
FT ACT_SITE 70
FT ACT_SITE 112
FT ACT_SITE 205
FT ACT_SITE 205
FT DISULFID 36
FT DISULFID 137
FT DISULFID 55
FT DISULFID 71
FT DISULFID 144
FT DISULFID 211
FT DISULFID 176
FT DISULFID 190
FT DISULFID 201
FT CARBOHYD 226
FT CARBOHYD 246
SQ SEQUENCE 253 AA; 27525 MW; 07EDB9F7 CRC32;

```

alignment_scores:

Quality: 274.50 Length: 143
Ratio: 2.745 Gaps: 3
Percent Similarity: 69.930 Percent Identity: 38.462

alignment block:

US-09-030-606-177 x SCCE_HUMAN

Align seg 1/1 to: SCCE_HUMAN from: 1 to: 253

```

7 TCGCAGCCCTGGAGCGGCGGAGTGTGATGGAAGAAACGAAATGTGTCTGTC 56
|||||
40 SerHisProThrPrlGlnAlaLeuTyrHisPheSerThrPheGlnCysG1 56
|||||
57 GGGCGTCTGGTGTGATCCGAGTGGTGGTGTGACCGGACACATGTTTCC 106
|||||
56 YglYValLeuValAsnGlnArgThrValLeuThrAlaAlaHisCysG1 73
|||||
107 AGAACCTCTACACATGGGGCTGGCGCTGACAGTCTTGAGCGGACCAA 156
|||||
73 etasnglutyThrValHisLeuGlySerAspThrLeu...GlyAspArg 88
|||||
157 GAGCCAGGAGGAGCAGATGTGTGAGCGGCGCTCTCCGTACGGCAGCAGA 206
|||||
89 Arg.....AlaGlnArgGlyLeuLysSerPheArgHisProG1 103
|||||
207 GTACACAGACCCCTGTGCTGCTAACGCTCATGCTCATGAGTTGAGC 256

```



```

103 YYYISERTHRLGINTHRLHISVALASNPSEULEUVALYSLEUASNS 120
257 AATCGGTGCCAGTGTGACACATCCGAGCATGACATTCGTTCGGAG 306
120 EGGHIALHRLGLEUSERSETMETVALYSVALRGLUEUPROSERARG 136
307 TCCCTACCCGCGGGGACTCTTCCTGCTTCGCTGGGCTGCTGCGC 356
137 CysGlnProProGlyThrThrCysThrValSerGlyTyrPheThr 153
357 GAACGATGCTGTG.....ATGCCATCCAGTCCAGCA 388
153 rserProAspValThrPheProSerAspLeuMetCysValAspVal 170
389 CTGTGGAGCGCTGGAGCTGTGAGAACCT 417
170 euIleSerProGlnAspCysThrLysVal 179

seq_name: SwissProt_37:KLK1_HUMAN

seq_documentation_block:
ID KLK1_HUMAN STANDARD; PRT; 262 AA.
AC P06870;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE GLANDULAR KALLIKREIN 1 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)
DE (KIDNEY/PANCREAS/SALIVARY GLAND KALLIKREIN).
GN KLK1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA.
CC PRIMATES; CARABRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PANCREAS;
RX MEDLINE; 86131629.
RA FUKUSHIMA D.; KITAMURA N.; NAKANISHI S.;
RT "Nucleotide sequence of cloned cDNA for human pancreatic kallikrein."
RL BIOCHEMISTRY 24:8037-8043(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY;
RX MEDLINE; 88269498.
RA EVANS B.A.; YUN Z.X.; CLOSE J.A.; TREGEAR G.W.; KITAMURA N.;
RA NAKANISHI S.; CALLEN D.F.; BAKER E.; HYLAND V.J.; SUTHERLAND G.R.;
RA RICHARDS R.I.;
RT "Structure and chromosomal localization of the human renal kallikrein gene."
RL BIOCHEMISTRY 27:3124-3129(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-SALIVARY GLAND;
RX MEDLINE; 90073574.
RA ANGERMANN A.; BERGMANN C.; APPELHANS H.;
RT "Cloning and expression of human salivary-gland kallikrein in Escherichia coli."
RL BIOCHEM. J. 262:787-793(1989).
RN [4]
RP SEQUENCE OF 17-262 FROM N.A.
RC TISSUE-KIDNEY;
RX MEDLINE; 86135264.
RA BAKER A.R.; SHINE J.;
RT "Human kidney kallikrein: cDNA cloning and sequence analysis."
RL DNA 4:445-450(1985).
RN [5]
RP SEQUENCE OF 25-262.
RC TISSUE-URINE;
RX MEDLINE; 89326688.
RA LU H.S.; LIN F.-K.; CHAO L.; CHAO J.;
RT "Human urinary kallikrein. Complete amino acid sequence and sites of glycosylation."
INT. J. PEPT. PROTEIN RES. 33:237-249(1989).

```

```

RN [6]
RP SEQUENCE OF 25-262, AND CARBOHYDRATE-BINDING SITES.
RC TISSUE-URINE;
RX MEDLINE; 88203586.
RA KELLERMANN J.; LOTTSPRECH F.; GEIGER R.; DEUTZMANN R.;
RT "Human urinary kallikrein -- amino acid sequence and carbohydrate attachment sites."
RL PROTEIN SEQ. DATA ANAL. 1:177-182(1988).
RN [7]
RP SEQUENCE OF 25-55.
RC TISSUE-URINE;
RX MEDLINE; 80114126.
RA LOTTSPRECH F.; GEIGER R.; HENSCHEN A.; KUTZBACH C.;
RT "N-terminal amino acid sequence of human urinary kallikrein homology with other serine proteases."
RL HOPE-SEYLER'S Z. PHYSIOL. CHEM. 360:1947-1950(1979).
RN [8]
RP SEQUENCE OF 28-47.
RC TISSUE-URINE;
RX MEDLINE; 86223893.
RA TAKAHASHI S.; IRIE A.; KATAYAMA Y.; ITO K.; MIYAKE Y.;
RT "N-terminal amino acid sequence of human urinary prokallikrein."
RL J. BIOCHEM. 99:989-992(1986).
CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS IN KININOMER TO RELEASE LYS-BRADYKININ.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-L-XAA BONDS IN SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE KALLIDIN (LYSL-BRADYKININ) FROM KININOMER INVOLVES HYDROLYSIS OF MET-L-XAA OR LEU-L-XAA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M25629; G186653; -.
DR EMBL; M33109; G186651; -.
DR EMBL; M33105; G186651; JOINED.
DR EMBL; M33106; G186651; JOINED.
DR EMBL; M33107; G186651; JOINED.
DR EMBL; M33108; G186651; JOINED.
DR EMBL; X13561; G34027; -.
DR PIR; A24696; KOHU.
DR MIM; 147910; -.
DR PROSITE; PS00134; TRYPSIN_HIS. 1.
DR PROSITE; PS00135; TRYPSIN_SER. 1.
DR PFAM; PF00089; trypsin. 1.
DR HSSP; P00757; ISGF.
KW HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN; MULTIGENE FAMILY; ZMOGEN; SIGNAL.
FT SIGNAL 1 18
FT PROPEP 19 24
FT CHARIN 25 262
FT ACT_SITE 65 65
FT ACT_SITE 120 120
FT ACT_SITE 214 214
FT CARBOHYD 93 93
FT CARBOHYD 102 102
FT CARBOHYD 104 104
FT CARBOHYD 108 108
FT CARBOHYD 165 165
FT CARBOHYD 167 167
FT DISULFID 31 174
FT DISULFID 50 66
FT DISULFID 153 220
FT DISULFID 185 199
FT DISULFID 210 235
FT VARIANT 145 145
PARTIAL (35% OF THE CHAINS).
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
Q -> E.

```

FT VARIANT 186 186 E -> K.
SQ SEQUENCE 262 AA; 28889 MW; 7B954AA7 CRC32;

alignment_scores:

Quality: 272.50 Length: 155
Ratio: 2.698 Gaps: 4
Percent Similarity: 65.161 Percent Identity: 38.710

alignment_block:

US-09-030-606-177 x KLK1_HUMAN ..

Align seg 1/1 to: KLK1_HUMAN from: 1 to: 262

```

4 CACTCGAGCCCTGGAGCGGCGGACGTGTCAGAAAAGATTGTTCTG 53
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 HisserginprotirglnAlaIaIeLeuThrHisPheSerThrPheGlnCy 50
54 CTGCGGCGTCTGTCGATCCGAGTGGGTGCTGTCACCCGACACTGTT 103
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
50 sGlyGlyIleLeuValHisArgGlnTrpValLeuThrAlaHisCysI 67
104 TCAGACACTCTACACCATCGGCGTGGCGCTGCGACAGCTTGAGCGGAC 153
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
67 IeSerAspAsnTrpGlnLeuTrpLeuGlyArgHisAsnLeu...PheAsp 82
154 CAGAGCGCAGGAGGAGGAGAGTGGAGCGGCGGCTCTCCGACGAGCAC 203
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
83 AspIuAsnThrAlaGlnPheValHisValSerGluSerPheProHisPr 99
204 AGACTACACAGACCCCTGCTGCTAAC..... 231
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
99 oGlyPheAsnMetSerLeuLeuGlnAsnHisThrArgGlnAlaAspGlu 116
232 .....GACCTCATGCTCATCAAGTGGAGAGATCCGTCGTCAG 270
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
116 sPlyrSerHisAspLeuMetLeuLeuArgLeuThrGlnProAlaAspThr 132
271 ...TCTGACACCATCCGAGCAGCATGCTGCTGTCAGTCCCTACGCG 317
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
133 IleThrAspAlaValIysValIaGlnLeuProThrGlnGlnProGlnIu 149
318 GGGGAACTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 357
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
149 IglYserThrCysLeuAlaSerGlyTrpGlySerIleGluProGlnAsp 166
358 .....AAGCATGCTGTGATGGCCATCCGACGCTCCGACACTGTGGAG 399
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
166 heSerPheProAspAspLeuGlnCysValAspLeuYsIleLeuProAsn 182
400 TGGGAGTGTGAGAG 414
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
183 AspIuGlyGlnIuYs 187

seq_name: SwissProt_37:KLK1_PAPHA

seq_documentation_block:
ID KLK1_PAPHA STANDARD: PRT; 258 AA.
AC 028773:
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE GLANDULAR KALLIKREIN 1 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)
DE (KIDNEY/PANCREAS/SALIVARY GLAND KALLIKREIN).
OS KLK1.
OS PAPIO HAMADRYAS (HAMADRYAS BABOON).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; PAPIO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PANCREAS;
RA PERELYGINA U.M.; KAMMERER C.M.; HENKEL R.D.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
```

```

CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-1-XAA BONDS IN
CC SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
CC KALLIDIN (LYSL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC MET-1-XAA OR LEU-1-XAA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: I43121; G871814; -.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PFAM: PF00089; trypsin; 1.
DR HSSP: P00752; 2KAT
KW HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN; PANCREAS;
KW ZYMOGEN; SIGNAL.
FT SIGNAL 1 18
FT PROPEP 19 24
FT CHAIN 25 258
FT ACT_SITE 62 62
FT ACT_SITE 117 117
FT ACT_SITE 210 210
FT DISULFID 31 170
FT DISULFID 47 63
FT DISULFID 149 216
FT DISULFID 181 195
FT DISULFID 206 231
FT DISULFID 266 291
FT CARBOHYD 90 90
FT CARBOHYD 99 99
FT CARBOHYD 101 101
FT CARBOHYD 105 105
FT CARBOHYD 161 161
FT CARBOHYD 161 161
SQ SEQUENCE 258 AA; 28339 MW; 0C5B883F CRC32;
```

alignment_scores:

Quality: 270.50 Length: 159
Ratio: 2.847 Gaps: 3
Percent Similarity: 59.748 Percent Identity: 37.107

alignment_block:

US-09-030-606-177 x KLK1_PAPHA ..

Align seg 1/1 to: KLK1_PAPHA from: 1 to: 258

```

7 TCGAGCGCTGGAGCGGCGGACGTGTCAGAAAAGATTGTTGCTC 56
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
32 SerGlnProTrpGlnAlaIaIaLeuThrHisPheSerThrPheGlnCysGl 48
57 GGGGCGTCTGTCGATCCGAGTGGGTGCTGTCACCCGACACTGTTCC 106
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
48 yGlyIleLeuValHisProGlnTrpValLeuThrAlaHisCysIleG 65
107 AGACTCTACACCATCGGCGTGGCGCTGCGACAGTCTTGAGCGGACCAA 156
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65 IysAspAsnTrpGlnLeuTrpLeuGlyArgHisAsnLeu...PheAsp 80
157 GAGCAGGAGCGGAGAGTGGAGCGGCGGCTCTCCGTCAGGACCCAGA 206
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
81 GluAspThrAlaGlnPheValHisValSerGluSerPheProHisProCy 97
207 GTACAACAGACCCCTGCTGCTAAC..... 231
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
97 sPheAsnMetSerLeuLeuYsAsnHisThrArgGlnAlaAspGluAspT 114
```

```

232 ..... GACCTCATGCTCATCAAGTTGGACGAATCCGTCGAGTCT 273
114 ysrerhisapleuemetleuleuargleuthrghlprohlauiether 130
274 GACACATCCGAGCATCGATGCTGTCGAGTCCGACCGCGGGGAA 323
131 aspalavalaGlnvalaGlnleuProthrGlnGlnProGlnvalaGlyse 147
324 CTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 348
147 thrGlyseuAlaSerGlyTTPeGlySerIleGluProGlnAsnPheserT 164
349 ..... CTGCTGGCGAAGCATGCT 366
164 yrProAspAspLeuGlnCysValaAspLeuGlyIleLeuProAsnAspLys 180
367 GGTGATGGCATCCAGTCCGAGACTGTG 393
181 CysAlaLysAlaAlaThrGlnLysVal 189

seq_name: SwissProt_37:PROS_HUMAN

seq_documentation_block:
ID   PROS_HUMAN          STANDARD:      PRT:      261 AA.
AC   P07288;
DT   01-APR-1988 (REL. 07, CREATED)
DI   01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   PROSTATE SPECIFIC ANTIGEN PRECURSOR (EC 3.4.21.77) (PSA) (GAMA-
GN   SEMINOPROTEIN) (SEMINOGLYCEIN) (SEMININ) (P-30 ANTIGEN).
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
RN   PROS_HUMAN
RP   TISSUE-PROSTATE;
RC   MEDLINE: 89183632.
RA   DIGBY M.R., ZHANG X.Y., RICHARD R.I.;
RT   "human prostate specific antigen (PSA) gene: structure and linkage to
RL   the kallikrein-like gene, hK2-1."
RN   NUCLEIC ACIDS RES. 17:2137-2137(1989).
[2]
RN   PROS_HUMAN
RP   MEDLINE: 89282407.
RA   KLOBECK H.-G., COMERARIO G., SCHULZ P., ARBUSOV V., FITTLER F.;
RT   "Genomic sequence of human prostate specific antigen (PSA).";
RL   NUCLEIC ACIDS RES. 17:3981-3981(1989).
[3]
RN   PROS_HUMAN
RP   SEQUENCE FROM N.A.
RC   TISSUE-LEUKOCYTE;
RC   MEDLINE: 89302050.
RA   LUNDWALL A.;
RT   "Characterization of the gene for prostate-specific antigen, a human
RT   glandular kallikrein."
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 161:1151-1159(1989).
[4]
RN   PROS_HUMAN
RP   SEQUENCE FROM N.A.
RC   TISSUE-PROSTATE;
RC   MEDLINE: 87190978.
RA   LUNDWALL A., LILJA H.;
RT   "Molecular cloning of human prostate specific antigen cDNA."
RL   FEBS LETT. 214:317-322(1987).
[5]
RN   PROS_HUMAN
RP   SEQUENCE FROM N.A.
RC   TISSUE-PROSTATE;
RC   MEDLINE: 89246551.
RA   HENTTU P., VIHKO P.;
RT   "cDNA coding for the entire human prostate specific antigen shows
RT   high homologies to the human tissue kallikrein genes."
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 160:903-910(1989).
[6]
RN   PROS_HUMAN
RP   SEQUENCE FROM N.A.

```

```

RC   TISSUE-PROSTATE;
RX   MEDLINE: 89165891.
RA   RIEGMAN P.H.J., VLEETSTRA R.J., VAN DER KORPUT J.A.G.M., ROMIJN J.C.,
RA   TRAPMAN J.;
RT   "Characterization of the prostate-specific antigen gene: a novel
RT   human kallikrein-like gene."
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 159:95-102(1989).
[7]
RN   PROS_HUMAN
RP   SEQUENCE FROM N.A.
RC   TISSUE-PROSTATE;
RA   MONNE M.M., MORENO J.M., MELE C.M., MUIHOLLAND G.M., COMELLA L.G.;
RL   SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
[8]
RN   PROS_HUMAN
RP   SEQUENCE OF 17-261 FROM N.A.
RX   MEDLINE: 88289366.
RA   SCHULZ P., STUCKA R., FELDMANN H., COMERARIO G., KLOBECK H.-G.,
RA   FITTLER F.;
RT   "Sequence of a cDNA clone encompassing the complete mature human
RT   prostate specific antigen (PSA) and an unspliced leader sequence."
RL   NUCLEIC ACIDS RES. 16:6226-6226(1988).
[9]
RN   PROS_HUMAN
RP   SEQUENCE OF 25-261.
RX   MEDLINE: 86205857.
RA   WATT K.W.K., LEE P.J., W'TINKULO T., CHAN W.P., LOOR R.;
RT   "Human prostate-specific antigen: structural and functional
RT   similarity with serine proteases."
RL   PROC. NATL. ACADE. SCI. U.S.A. 83:3166-3170(1986).
[10]
RN   PROS_HUMAN
RP   SEQUENCE OF 25-261.
RX   MEDLINE: 88082806.
RA   SCHALLER J., AKIYAMA K., TSUDA R., HARA M., MARTI T., RICKLI E.E.;
RT   "Isolation, characterization and amino-acid sequence of gamma-
RT   semioprotein, a glycoprotein from human seminal plasma."
RL   EUR. J. BIOCHEM. 170:111-120(1987).
[11]
RN   PROS_HUMAN
RP   3D-STRUCTURE MODELLING.
RX   MEDLINE: 95218633.
RA   VILLOUTREIX B.O., GETZOFF E.D., GRIFFIN J.H.;
RT   "A structural model for the prostate disease marker, human prostate-
RT   specific antigen."
RL   PROTEIN SCI. 3:2033-2044(1994).
[12]
RN   PROS_HUMAN
RP   VESICLE PROTEIN THUS LEADING TO THE LIQUEFACTION OF THE SEMINAL
CC   COAGULUM.
CC   -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: TYR-I-XAA.
CC   -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC   TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
-----
CC   THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
-----
DR   EMBL: X14810; G296671; -
DR   EMBL: X13940; G35723; -
DR   EMBL: X13941; E14018; ALT_SEQ.
DR   EMBL: X13942; E14019; -
DR   EMBL: X13943; E14020; -
DR   EMBL: X13944; E14022; -
DR   EMBL: X05332; G35741; -
DR   EMBL: X07730; -; NOT_ANNOTATED_CDS.
DR   EMBL: M27274; G190553; -
DR   EMBL: M26663; G618464; -
DR   EMBL: U17040; G595946; -
DR   EMBL: M24543; G511857; ALT_SEQ.
DR   PIR: S03604; S03604; -
DR   PIR: A26757; A26757; -
DR   PIR: A32297; A32297; -
DR   PIR: A32423; A32423; -
DR   PDB: 1PFA; 26-JAN-95.

```

DR MIM: 176820; .
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PFAM: PF00089; trypsin; 1.
 KM HYDROLASE; SERINE PROTEASE; GLYCOPROTEIN; ANTIGEN; ZYMOGEN; SIGNAL;
 KW 3D-STRUCTURE.
 FT SIGNAL 1 18 POTENTIAL.
 FT PROPEP 19 24 ACTIVATION PEPTIDE.
 FT CHAIN 25 261 PROSTATE SPECIFIC ANTIGEN.
 FT ACT_SITE 120 65 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 69 69
 FT DISULFID 31 173 BY SIMILARITY.
 FT DISULFID 50 66 BY SIMILARITY.
 FT DISULFID 152 219 BY SIMILARITY.
 FT DISULFID 184 198 BY SIMILARITY.
 FT DISULFID 209 234 BY SIMILARITY.
 FT CONFLICT 64 64 A -> T (IN REF. 8).
 FT CONFLICT 73 73 T -> I (IN REF. 5).
 FT CONFLICT 86 86 I -> T (IN REF. 5).
 FT CONFLICT 94 94 H -> T (IN REF. 9).
 FT CONFLICT 136 136 V -> M (IN REF. 8).
 FT CONFLICT 165 168 FLTP -> HLTDQM (IN REF. 9).
 FT CONFLICT 175 175 D -> Q (IN REF. 9).
 FT CONFLICT 175 175 D -> P (IN REF. 5).
 FT CONFLICT 184 184 C -> Q (IN REF. 5).
 FT CONFLICT 260 260 N -> D (IN REF. 5).
 SO SEQUENCE 261 AA: 28741 MW: 4723552E CRC32;

alignment_scores: Quality: 270.00 Length: 170
 Ratio: 2.700 Gaps: 4
 Percent Similarity: 58.824 Percent Identity: 34.706

alignment_block:

US-09-030-606-177 x PROS_HUMAN

Align seg 1/1 to: PROS_HUMAN from: 1 to: 261

4 CACITCGACCCCTGCGAGCGGACGTGTCATGGAAGAAAGATGTTCTG 53
 ||||||||||||||||||| : : : : : : : : : : : : : : : :
 34 HisserginProtrpglnValleuValalaserArgylarGalaValcy 50
 54 CTGGGGGCTGCTGGTCATCGGCGAGTGGTGTCTGCACCCGACACTTT 103
 | :
 50 SGlyGlyValleuValHisProglntPrValleuthrAlaHisCysI 67
 104 TCAGAACTCCTACACCATCGGCTGGGCTGCACAGTCTTGAGGCGGAC 153
 :
 67 leaVasnlrSerVallleuLeuenglArgHIsSerleuPheHispro 83
 154 CAAGAGCCAGGAGGACGATGTTGAGCGGCTCTCGGACGCGACCC 203
 :
 84 GlusprHrNgly...GlnValrPheglInValSerHisSerPheProHIspr 99
 204 AGAGTACACAGACCCCTGCTGCTGCTAAC..... 231
 | :
 99 OleutyrAspMetSerleuLeuLysAsnAlrPheleuArgProgluysara 116
 232GACCTCATGCTCATCAGATTGACGATCGGTGCGAG 270
 ||||||||||||||||||| : : : : : : : : : : : : : : : :
 116 spserSerHisAspMetleuLeuArgleuSerGlnProAlaGluLeu 132
 271 TGTGACACCATCGGAGCATGACATTGCTTGGACGCGCTTACCGCGGG 320
 :
 133 ThrspAlaValLysValMetAspLeuProthrGlnGluProAlaLeuGl 149
 321 GAACCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 348
 | :
 149 YThrThrcysTyrAlaSerGlyTyrGlySerlleGluProGluGluPheL 166

349CTGCTGGCAGACAT 363
 166 eutHrProLysLysleuGlnCysValaAspMetHIsValleuSerAsnAsp 182
 364 GCTGTGATTCATCCAGTCAGTCCAGACTGTG..... 393
 :
 183 ValCysAlaGlnValHisProGlnLysValThrLysPheMetLeuCysAl 199
 394 ..GAGGCTGG 402
 :
 199 agLysArgTyr 202
 seq_name: SwissProt_37:Klk2_HUMAN
 seq_documentation_block:
 ID Klk2_HUMAN STANDARD; PRT; 261 AA.
 AC P20151;
 DT 01-FEB-1991 (REL. 17, CREATED)
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE GLANDULAR KALLIKREIN 2 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)
 DE (PROSTATE) (HGK-1).
 GN KLK2.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 88054467.
 RA SCHIEDLICH L.J., BENNETTS B.H., MORRIS B.J.;
 RL "Primary structure of a human glandular kallikrein gene.";
 RN DNA 6:429-437(1987).
 [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=PROSTATE;
 RC MEDLINE: 92324494.
 RA RIEGMAN P.H., VLEETSTRA R.J., DER KOPPUT H.A., ROMIJN J.C.,
 RA TRAPMAN J.;
 RT "Identification and androgen-regulated expression of two major human
 glandular kallikrein-1 (hgk-1) mRNA species.";
 RL MOL. CELL. ENDOCRINOL. 76:181-190(1991).
 CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
 IN KININOGEN TO RELEASE LYS-BRADYKININ.
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-1-XAA BONDS IN
 SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
 KALLIDIN (LYSIL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLISIS OF
 MET-1-XAA OR LEU-1-XAA.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sb.ch/announce/>
 CC or send an email to license@isb-sb.ch).
 CC
 DR EMBL: M18156; -; NOT_ANNOTATED_CDS.
 DR EMBL: M18157; G386842; -.
 DR EMBL: S39329; E79223; -.
 DR PIR: A29586; A29586.
 DR MIM: 147960; -.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PFAM: PF00089; trypsin; 1.
 DR HSSP: P07288; 1PFA.
 KM HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN;
 KW MULTIGENE FAMILY; ZYMOGEN; SIGNAL.
 FT SIGNAL 1 18 PROBABLE.
 FT PROPEP 19 24 ACTIVATION PEPTIDE (PROBABLE).
 FT CHAIN 25 261 GLANDULAR KALLIKREIN 2.
 FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.

FT ACT_SITE 120 120 CHARGE RELAY SYSTEM.
 FT ACT_SITE 213 213 CHARGE RELAY SYSTEM.
 FT DISULFID 31 173 BY SIMILARITY.
 FT DISULFID 50 66 BY SIMILARITY.
 FT DISULFID 152 219 BY SIMILARITY.
 FT DISULFID 184 198 BY SIMILARITY.
 FT DISULFID 209 234 BY SIMILARITY.
 FT CARBOHYD 102 102 POTENTIAL.
 SO SEQUENCE 261 AA; 28671 MW; 99CF6275 CRC32;

alignment_scores:
 Quality: 266.50 Length: 160
 Ratio: 2.835 Gaps: 3
 Percent Similarity: 58.750 Percent Identity: 36.250

alignment_block:

US-09-030-606-177 x KLR2_HUMAN ..

Align seg 1/1 to: KLR2_HUMAN from: 1 to: 261

```

4 CACTGCGACCCCTGCGAGCGGCGACTGTCATGAGAAAGCAATTGTCTG 53
|||||
34 HISSerGlnProTyrGlnValAlaValTyrSerHisGlyTyrPalaHisCy 50
CTCGGGCGCTGTCGATCCGCGAGTGGTGGTCGACCGCGACACTGT 103
|||||
50 SGIyGlyValLeuValHisProGlnTyrPalaLeuThrAlaAlaHisCy 67
|||||
104 TCAGAACTCTTACACATCGCGGCTGGCTGCGACAGTGTAGCGCGAC 153
|||||
67 EulYLyAsnSerGlnValTyrLeuGlyArgHisAsnLeuPheGluPro 83
|||||
154 CAGAGCCAGGAGCCAGATGTCGAGCGGCGCTCTCGTCAGCGACCC 203
|||||
84 GluaPrlngly...GlnArgValProValSerHisSerPheProHisPr 99
|||||
204 AGAGTACAACAGACCTTGGCTGCTC..... 225
|||||
99 OleuTyrAsnMetSerLeuLeuHisGlnSerLeuArgProAsnGluA 116
|||||
226 .....GCTAACAGCCTCATGTCATCAAGTGGACAGTCCGTCGCGAG 270
|||||
116 spSerSerHisAsnLeuMetLeuLeuArgLeuSerGlnProAlaLysIle 132
|||||
271 TCTGACACCATCCGAGCATGACATGCTTGCAGTGGCTGACCGCGGG 320
|||||
133 ThrAspValAllyValLeuGlyLeuProThrGlnGluProAlaLeuGl 149
|||||
321 GAACCTTGGCTGCTTCTGCTGGGGT..... 348
|||||
149 YThrTyrCysTyrAlaSerGlyTyrPoleuSerTleuGluProGluGluPhe 166
|||||
349 .....CTGCTGGCGAAGCAT 363
|||||
166 eulArgProArgSerLeuGlnCysValSerLeuHisLeuLeuSerAsnAsp 182
|||||
364 GCTGTGATTCATCCATCCAGTCCGAGACTGTG 393
|||||
183 MetCysAlaArgAlaTyrSerGluLysVal 192
|||||
seq_name: SwissProt_37:PROS_MACMU

```

```

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; CERCOPITHECIDAE; CERCOPITHECINAE; MACACA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93363642.
RA GAUTHIER E.R., CHADELAINE P., TREMBLAY R.R., DUBE J.Y.;
"Characterization of rhesus monkey prostate specific antigen cDNA.";
RL BIOCHIM. BIOPHYS. ACTA 1174:207-210(1993).
CC -1- FUNCTION: PSA PRESUMABLY HYDROLYZE THE HIGH MOLECULAR MASS
SEMINAL VESICLE PROTEIN THUS LEADING TO THE LIQEFRACTION OF
THE SEMINAL COAGULUM. PSA IS A GLANDULAR KALLIKREIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X73560; G311844; -.
CC PIR; S34239; S34239.
CC PIR; S35711; S35711.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC PRAM; PF00089; trypsin; 1.
CC HSSP; P07288; 1PFA.
CC KW HYDROLASE; SERINE PROTEASE; GLYCOPROTEIN; ANTIGEN; ZMOGEN; SIGNAL.
FT SIGNAL 1 18
FT PROPEP 19 24
FT CHAIN 25 261
FT ACT_SITE 65 65 PROSTATE SPECIFIC ANTIGEN.
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 31 173 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 50 66 BY SIMILARITY.
FT DISULFID 152 219 BY SIMILARITY.
FT DISULFID 184 198 BY SIMILARITY.
FT DISULFID 209 234 BY SIMILARITY.
FT CARBOHYD 102 102 POTENTIAL.
SO SEQUENCE 261 AA; 28816 MW; 0B73118F CRC32.

```

alignment_scores:
 Quality: 262.00 Length: 170
 Ratio: 2.646 Gaps: 4
 Percent Similarity: 58.235 Percent Identity: 33.529

alignment_block:

US-09-030-606-177 x PROS_MACMU ..

Align seg 1/1 to: PROS_MACMU from: 1 to: 261

```

4 CACTGCGACCCCTGCGAGCGGCGACTGTCATGAGAAAGCAATTGTCTG 53
|||||
34 HISSerGlnProTyrGlnValAlaValTyrSerHisGlyTyrPalaHisCy 50
CTCGGGCGCTGTCGATCCGCGAGTGGTGGTCGACCGCGACACTGT 103
|||||
50 SGIyGlyValLeuValHisProGlnTyrPalaLeuThrAlaAlaHisCy 67
|||||
104 TCAGAACTCTTACACATCGCGGCTGGCTGCGACAGTGTGAGCGCGAC 153
|||||
67 leArgSerAsnSerValIleLeuLeuGlyArgHisAsnProTyrTyrPro 83
|||||
154 CAGAGCCAGGAGCCAGATGTCGAGCGGCGCTCTCGTCAGCGACCC 203
|||||
84 GluaPrlngly...GlnArgValPheGlnValSerHisSerPheProHisPr 99
|||||
204 AGAGTACAACAGACCTTGGCTGCTGCTAAC..... 231
|||||

```

```

99 oLeuTyraMetSerLeuLeuYsaAsnArgTyLeuGlyProGlyAsp 116
232 .....GACCTCATGCTCATCAAGTTGACGCAATCCGCTCGAG 270
116 spSerSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaGluIle 132
271 TCAGACACATCCGAGCATCAGCATGCTTCGACGACCCCTACCGGGG 320
133 ThrAspAlaValGlnValIleAspLeuProThrProIleProGluLeuG 149
321 GAACCTCTGCTGCTTCTGCTGGGT..... 348
149 yThrThyCysTyraSerGlyArgGlySerIleGluProGluGluHis 166
349 .....CTGCTGGCGAAGAT 363
166 eUThrProLysLysLeuGlnCysValAspLeuHisIleSerAsnAsp 182
364 GCTGTGATTGCTCCATCCAGTCCAGACTGTG..... 393
183 ValCysAlaGlnValHisSerGlnLysValThrLysPheMetLeuCysAl 199
394 .GGAGCTGG 402
199 agLysSerTrp 202

seq_name: swissprot_37:KLKL_MOUSE

seq_documentation_block:
ID KLKL_MOUSE STANDARD: PRT; 259 AA.
AC M15948;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DE 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE GLANDULAR KALLIKREIN K22 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)
DE (MGK-22) (EPIDERMAL GROWTH FACTOR-BINDING PROTEIN TYPE A) (EGF-BP A)
DE (NERVE GROWTH FACTOR BETA CHAIN ENDOPEPTIDASE) (BETA-NGF-
DE ENDOPEPTIDASE).
DE KLR22 OR KLR-22.
GN MGS MUSCULUS (MOUSE).
OS EUKARIOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIROGNATHI; MURIDAE; MURINAE; MGS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-SALIVARY GLAND;
RX MEDLINE: 88107594.
RA DRINKWATER C.C., EVANS B.A., RICHARDS R.I.;
RA "Mouse glandular kallikrein genes: identification and
RA characterization of the genes encoding the epidermal growth factor
RA binding proteins.";
RT BIOCHEMISTRY 26:6750-6756(1987).
RN [2]
RP SEQUENCE OF 25-54.
RX MEDLINE: 91190897.
RA FAHNESTOCK M., WOO J.E., LOPEZ G.A., SNOW J., WALZ D.A., ARICI M.J.,
RA MOBLEY W.C.;
RA "Beta-NGF-endopeptidase: structure and activity of a kallikrein
RA encoded by the gene mgk-22.";
RT BIOCHEMISTRY 30:3443-3450(1991).
RN [3]
RP SEQUENCE OF 25-41.
RC TISSUE-SUBMANDIBULAR GLAND;
RX MEDLINE: 92348355.
RA PETERS J., TAKAHASHI S., TADA M., MIYAKE Y.;
RA "MGK-6-derived true tissue kallikrein is synthesized, processed, and
RA targeted through a regulated secretory pathway in mouse pituitary
RA AtT-20 cells.";
RT J. BIOCHEM. 111:643-648(1992).
RN [4]
RP SEQUENCE OF 17-54 AND 70-120 FROM N.A.
RX MEDLINE: 87250386.
RA EVANS B.A., DRINKWATER C.C., RICHARDS R.I.;
RA "Mouse glandular kallikrein genes. Structure and partial sequence

```

```

RT analysis of the kallikrein gene locus."
J. BIOL. CHEM. 262:8027-8034(1987).
CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-1-XNA BONDS IN
CC SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
CC KALLIDIN (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC MET-1-XNA OR LEU-1-XNA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ALSO KNOWN AS THE
CC TRYPsin FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M17979; G387167; ALT SEQ.
DR EMBL: M17977; G387167; JOINED.
DR EMBL: M17978; G387167; JOINED.
DR EMBL: M18598; G198560; -.
DR EMBL: M18618; G198561; -.
DR PIR: A29746; A29746.
DR PIR: A38356; A38356.
DR PROSITE: PS00134; TRYPsin_HIS; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
DR PFM: PF00089; trypsin; 1.
DR HSSP: P00757; 1SGF.
KW HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN;
KW MULTIGENE FAMILY; ZYMOGEN; SIGNAL.
FT SIGNAL 1 17
FT PROPEP 18 24 ACTIVATION PEPTIDE.
FT CHAIN 25 259 GLANDULAR KALLIKREIN K22.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.
FT ACT_SITE 118 118 CHARGE RELAY SYSTEM.
FT ACT_SITE 211 211 CHARGE RELAY SYSTEM.
FT ACT_SITE 31 171 BY SIMILARITY.
FT DISULFID 50 66 BY SIMILARITY.
FT DISULFID 150 217 BY SIMILARITY.
FT DISULFID 182 196 BY SIMILARITY.
FT DISULFID 207 232 BY SIMILARITY.
FT CARBOHYD 102 102 PROBABLE.
SQ SEQUENCE 259 AA; 28384 MW; 3802165A CRC32;

alignment_scores:
Quality: 260.00 Length: 180
Ratio: 2.342 Gaps: 6
Percent Similarity: 61.667 Percent Identity: 33.889

alignment_block:
US-09-030-606-177 x KLKL_MOUSE ..
Align seg 1/1 to: KLKL_MOUSE from: 1 to: 259

4 CACTCGACGCCCTGGGACAGCGGCGACTGGTCATGAAACGAAATGTTCTG 53
:::|||||:::|||||:::|||||:::|||||:::|||||
34 AsnSerGlnProThrGlnValAlaValAlaValTyTyLeuAspIleuTyLeuGly 50
:::|||||:::|||||:::|||||:::|||||:::|||||
54 CTGGGCGCTCTGCTGATCCGAGTGGGCGTGTACGCCGACATGTT 103
:::|||||:::|||||:::|||||:::|||||:::|||||
50 sGlyGlyValLeuLeuAspArgAsnTrpValLeuThrAlaAlaHisCysT 67
:::|||||:::|||||:::|||||:::|||||:::|||||
104 TCAGACATCTCTACACATCGGCGCTGGCTGCACAGTCTTGAGCGGAC 153
:::|||||:::|||||:::|||||:::|||||:::|||||
67 yGlnAspLysTyraHisIleTrpLeuGlyLysAsnLysLeuPheGlnAsp 83
:::|||||:::|||||:::|||||:::|||||:::|||||
154 CAAGAGCCAGGAGGACGATGTGAGAGCGGCGCTCTCGTACGCGACCC 203
:::|||||:::|||||:::|||||:::|||||:::|||||
84 ...GluProSerAlaGlnHisArgLeuValSerLysSerPheProHisPr 99

```

```

204 AGAGTACACAGACCCCTG.....CTCG 226
    |||
    |||
99  oasphensmetserleuInserValProthrGlyAlaaspleus 116
227 CTACGACCTATGCTATCAAGTTGAGCAATCCGTCCGAGTCTGAC 276
    |||
    |||
116  erAsnAspleuMetleuThrGlySerLysProAlaAspleuThrAsp 132
277 ACATCCGAGCATGATGCTTCGAGTCCCTACCGGGGAGACTC 326
    |||
    |||
133 VALValLysProIleAspleuProThrGlyProLysLeuGlySerTh 149
327 TTGCCTCTTCTGCTGCTGCTGCTGCGACAGATCGATGATGCA 376
    |||
    |||
149  rCysLeuAlaSerGlyTyrPglySerIleAsnGlnLeuIle..... 162
377 TCAGACTCCAGACGTGTGGAGGCTGGAGGTGTGAGAACCTTCCCAACC 426
    |||
    |||
163  .....TyrGlnAsnPr 166
427 TGGCAGGTTGTACCATTTGGCAACTTCAGATCGAGAGAGCTCTGCTG 476
    |||
    |||
166  o.....AsnAspleu.GlnCysValSerIleLysLeu 176
477 CATCTCACTGGGTGCTCACTCACTCACTCACTGCATC 512
    |||
    |||
177 HtSProAsnGlnValCysValLysAlaHtStIleu 188

seq_name: SwissProt_37:KLK3_MOUSE

seq_documentation_block:
ID  KLK3_MOUSE  STANDARD;  PRT;  261 AA.
AC  P00756;
DT  21-JUL-1986 (REL. 01, CREATED)
DT  21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT  15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE  GLANDULAR KALLIKREIN K3 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)
DE  (MGK-3) (7S NERVE GROWTH FACTOR GAMMA CHAIN) (GAMMA-NGF).
GN  KLK3 OR KLK-3 OR NGFG.
OS  MUS MUSCULUS (MUSE).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC  RODENTIA; SCIDROGNATHI; MURIDAE; MURINAE; MUS.
NC  [1]
RP  SEQUENCE FROM N.A.
RA  DILRICH A., GRAY A., WOOD W.I., HAYFLICK J., SEEBURG P.H.;
RT  "Isolation of a cDNA clone coding for the gamma-subunit of mouse
RL  nerve growth factor using a high-stringency selection procedure.";
RN  DNA 3:387-392(1984).
RP  SEQUENCE FROM N.A.
RA  MEDLINE; 8537431.
RA  EVANS B.A., RICHARDS R.I.;
RT  "Genes for the alpha and gamma subunits of mouse nerve growth factor
RT  are contiguous.";
RL  EMBO J. 4:133-138(1985).
RN  [3]
RP  SEQUENCE OF 25-261.
RX  MEDLINE; 81264363.
RA  THOMAS K.A., BAGLAN N.C., BRADSHAW R.A.;
RT  "The amino acid sequence of the gamma-subunit of mouse submaxillary
RT  gland 7 S nerve growth factor.";
RL  J. BIOL. CHEM. 256:9156-9166(1981).
RN  [4]
RP  X-RAY CRYSTALLOGRAPHY (3.15 ANGSTROMS) OF 7S COMPLEX.
RC  STRAIN-SWISS WEBSTER; TISSUE-SUBMAXILLARY GLAND;
RX  MEDLINE; 98035451.
RA  BAX B., BLUNDELL T.L., MURRAY-RUST J., MCDONALD N.O.;
RT  "Structure of mouse 7S NGF: a complex of nerve growth factor with
RT  four binding proteins.";
CC  STRUCTURE 5:1275-1285(1997).
CC  FUNCTION: 7S NGF ALPHA CHAIN STABILIZES THE 7S COMPLEX. THE BETA
CC  DIMER PROMOTES NEURITE GROWTH. THE GAMMA CHAIN IS AN ANGININE-

```

```

CC  SPECIFIC PROTEASE; IT MAY ALSO HAVE PLASMINOGEN ACTIVATOR
CC  ACTIVITY, AS WELL AS MITOGENIC ACTIVITY FOR CHICK EMBRYO
CC  FIBROBLASTS.
CC  -1- SUBUNIT: 7S NERVE GROWTH FACTOR IS COMPOSED OF TWO ALPHA CHAINS,
CC  A BETA DIMER COMPOSED OF IDENTICAL CHAINS, AND TWO GAMMA CHAINS.
CC  -1- THIS PRECURSOR IS CLEAVED INTO SEGMENTS TO PRODUCE THE ACTIVE
CC  FORM OF THE GAMMA CHAIN, WHICH OCCURS NATURALLY AS COMBINATIONS
CC  OF EITHER TWO OR THREE SEGMENTS HELD TOGETHER BY DISULFIDE
CC  BONDS: B1 + A OR B1 + C + B2.
CC  -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC  TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC  -----
CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch).
CC  -----
DR  EMBL; X01389; G53374; -
DR  EMBL; X01798; G582429; -
DR  EMBL; X01799; E73575; -
DR  PIR; A00942; NMSG.
DR  PDB; 1SGF; 27-MAY-98.
DR  MGD; MGI:97322; NGFG.
DR  PROSITE; PS00134; TRYPSIN_HIS; 1.
DR  PROSITE; PS00135; TRYPSIN_SER; 1.
DR  PFM; PF00089; trypsin; 1.
DR  KM HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN;
DR  MULTIGENE FAMILY; ZMOGEN; SIGNAL; GROWTH FACTOR; 3D-STRUCTURE.
FT  SIGMAP 1 18 PROBABLE.
FT  PROPEP 1 24 ACTIVATION PEPTIDE.
FT  CHAIN 25 261 GLANDULAR KALLIKREIN K3.
FT  CHAIN 25 107 NERVE GROWTH FACTOR GAMMA CHAIN 1.
FT  ACT_SITE 112 261 NERVE GROWTH FACTOR GAMMA CHAIN 2.
FT  ACT_SITE 65 65 CHARGE RELAY SYSTEM.
FT  ACT_SITE 120 120 CHARGE RELAY SYSTEM.
FT  ACT_SITE 213 213 CHARGE RELAY SYSTEM.
FT  DISULFID 31 173 PROBABLE.
FT  DISULFID 50 66 GLANDULAR KALLIKREIN K3.
FT  DISULFID 152 219 NERVE GROWTH FACTOR GAMMA CHAIN 1.
FT  DISULFID 184 198 NERVE GROWTH FACTOR GAMMA CHAIN 2.
FT  DISULFID 209 234 CHARGE RELAY SYSTEM.
FT  CARBOHYD 102 102 CHARGE RELAY SYSTEM.
FT  DOMAIN 25 107 SEGMENT B1.
FT  DOMAIN 112 261 SEGMENT A.
FT  DOMAIN 112 164 SEGMENT C.
FT  DOMAIN 165 261 SEGMENT B2.
FT  CONFLICT 108 111 MISSING (IN REF. 2).
SQ  SEQUENCE 261 AA; 28998 MW; 7B9C5256 CRC32;

alignment_scores:
Quality: 255.00 Length: 154
Ratio: 2.476 Gaps: 3
Percent Similarity: 66.883 Percent Identity: 32.468

alignment_block:
US-09-030-606-177 x KLK3_MOUSE ..
Align seg 1/1 to: KLK3_MOUSE from: 1 to: 261

4 CACTCGAGCCCTGCGAGCGGCGGCTGTCATGGAACGATTTGCTG 53
|||||
34 AsnserGlnProThrPhtValAlaValTyrArgTyrThrGlnTyrLeu 50
|||||
54 CTGCGGGCTGTGTCATCGGAGGTGGTCTGACGGCAGACTGTT 103
|||||
50 sGlyValLeuLeuAspProAsnTyrValLeuThrAlaAlaHtScys 67
|||||
104 TCAGACTCTACACATGCGGCTGGCGCTGCACACTTTCAGCGGAC 153
|||||

```



```

4 CACTGCGAGCCCTGGCAGCGGCGAGCTGATGAGAAAGCATTTGTCG 53
   :::::::::::::::::::::::::::: ||||| ::::|
19 AsnSerGlnProTPrpInValAlaIleIle.....AsnGluTyrLeu 33
54 CTCGGGCGCTCTGTGATCCGAGTGGGTGCTGCAGCCGACACTGTT 103
   :::::::::::::::::::: ||::::::::::|
33 sGlyGlyValLeuIleAspProSerTPrpValIleThrAlaIleHisCys 50
104 TCAGAACTCTACACACATCGGGCTGGGCTGCACAGTCTTGAGCCGAC 153
   ::::::::::| | | | | ::::::::::| | |
50 YrSerAsnTyrTyrHisValLeuLeuGlyArgAsnAsnLeuHegLysp 66
154 CAAGAGCCAGGAGGAGCAGATGTTGGAGGCGCCTTCGTCAGGCGACC 203
   ||||| ::::| | | | | ::::| | |
67 ..GluProPheAlaGlnTyrArgPheValAsnGlnSerPheProHisP 82
204 AGAGTACACAGACACCTGCTC..... 225
   ::::| | | | |
82 AspTyr...LysProPheLeuMetArgAsnHisThrArgGlnArgGly 98
226 .....GCTACAGACTCATGCTCATCAAGTGGAGCAATCCGTGCC 267
   ::::::::::| | | | | ::::| | |
98 sPAspTyrSerAsnAspLeuMetLeuLeuHisLeuSerGluProAlaAsp 114
268 GAGTGTGACACCATCCGAGCATGATGCTGTCGAGCGCTACCGC 317
   ::::| | | | | ::::| | |
115 IleThrAspGlyValIleValIleAspLeuProThrGluGluProLys 131
318 GGGGAACTCTGCTCTGCTTCTGCTGGGGT..... 348
   ::::::::::| | | | | ::::| | |
131 lGlySerThrCysLeuAlaSerGlyTyrPheLysSerThrLysProLeuAsn 148
349 .....CTGCTGGCGAAG 360
   ::::::::::| | | | |
148 rPgIleuProAspAspLeuGlnCysValAsnHisLeuLeuSerAsn 164
361 GATGCTGTATGATCCATCCAGTCCAGACTG 393
   ::::| | | | |
165 GluLysCysIleGluAlaTyrGlnGlnLysVal 175

seq_name: SwissProt_37:KLKA_MOUSE

seq_documentation_block:
ID KLKA_MOUSE STANDARD; PRT; 261 AA.
AC P15946;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE GLANDULAR KALLIKREIN K11 PRECURSOR (EC 3.4.21.35), SUBMANDIBULAR
DE (TISSUE KALLIKREIN) (MKG-11).
OS KUKU1 OR KUK-11.
GN MOS MUSCULUS (MUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIROGNATHI; MORIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-LIVER;
RX MEDLINE; 89083511.
RA DRINKWATER C.C.; RICHARDS R.I.;
RT "sequence of mkg-11, a mouse glandular kallikrein gene.";
RL NUCLEIC ACIDS RES. 16:10918-10918(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85257431.
RA EVANS B.A.; RICHARDS R.I.;
RT "genes for the alpha and gamma subunits of mouse nerve growth factor
are contiguous.";
RL EMBO J. 4:133-138(1985).
RN [3]
RP SEQUENCE OF 16-54 AND 69-122 FROM N.A.
RX MEDLINE; 87250386.
RA EVANS B.A.; DRINKWATER C.C.; RICHARDS R.I.;

```

```

RT "Mouse glandular kallikrein genes. Structure and partial sequence
RT analysis of the kallikrein gene locus.";
RL J. BIOL. CHEM. 262:8027-8034(1987).
CC -1- FUNCTION: GLANDULAR KALLIKREIN CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-1-XAA BONDS IN
CC SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
CC KALLIDIN (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC MET-1-XAA OR LEU-1-XAA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC
DR EMBL; X13215; G53066; -
DR EMBL; X13216; G582395; ALT_SEQ.
DR EMBL; X13217; E14069; -
DR EMBL; X13218; E14070; -
DR EMBL; X13218; E7100; -
DR EMBL; M18590; G554179; ALT_SEQ.
DR EMBL; M18610; G198537; -
DR PIR; S01971; S01971.
DR MGD; MGI:892023; KLK11.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
DR PFAM; PF00089; trypsin; 1.
DR HSSP; P00757; ISGF.
DR KW HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN;
KW MULTIGENE FAMILY; ZYMOGEN; SIGNAL.
FT SIGNAL 1 18
FT PROPEP 19 24 ACTIVATION PEPTIDE (PROBABLE).
FT CHAIN 25 261 GLANDULAR KALLIKREIN K11.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM.
FT ACT_SITE 213 213 CHARGE RELAY SYSTEM.
FT DISULFID 31 173 BY SIMILARITY.
FT DISULFID 50 66 BY SIMILARITY.
FT DISULFID 152 219 BY SIMILARITY.
FT DISULFID 184 198 BY SIMILARITY.
FT DISULFID 209 234 BY SIMILARITY.
FT CARBOHYD 102 102 PROBABLE.
SQ SEQUENCE 261 AA; 28727 MW; FD904C22 CRC32;

alignment_scores:
Quality: 250.50 Length: 146
Ratio: 2.665 Gaps: 3
Percent Similarity: 64.384 Percent Identity: 36.301

alignment_block:
US-09-030-606-177 x KLKA_MOUSE
Align seg 1/1 to: KLKA_MOUSE from: 1 to: 261

4 CACTGCGAGCCCTGGCAGCGGCGAGCTGATGAGAAAGCATTTGTCG 53
   :::::::::::::::::::::::::::: ||||| ::::|
34 AsnSerGlnProTPrpInValAlaIleIle.....AsnGluTyrLeu 50
54 CTCGGGCGCTCTGTGATCCGAGTGGGTGCTGCAGCCGACACTGTT 103
   :::::::::::::::::::: ||::::::::::|
50 sGlyGlyValLeuIleAspProSerTPrpValIleThrAlaIleHisCys 67
104 TCAGAACTCTACACACATCGGGCTGGGCTGCACAGTCTTGAGCCGAC 153
   ::::::::::| | | | | ::::::::::| | |
67 lSvalSerGlnTyrAsnValTyrLeuGlyLysThrLysLeu...PheGln 82
154 CAAGAGCCAGGAGGAGCAGATGTTGGAGGCGCCTTCGTCAGGCGACC 203

```



```

      131 SERCSPFQIALALAGLYTHGLncysLeuilesergLyTPGly..... 145
      146 .....AsnThcInserilegly 151

seq_name: SwissProt_37:KLK2_RAT

seq_documentation_block:
ID   KLK2_RAT          STANDARD;             PRT;           259 AA.
AC   P00759;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE) ~
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   TONIN PRECURSOR (EC 3.4.21.35) (ESTERASE 1) (S2 KALLIKREIN) (RGK-2)
DE   (RGKG-5).
DN   KLK2 OR KLK-2 OR TON.
OS   RATUUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MORIDAE; MORINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86051477.
RA   ASHLEY P.L., MCPDONALD R.J.;
RT   "Kallikrein-related MNAS of the rat submaxillary gland: nucleotide
RT   sequences of four distinct types including tonin.";
RL   BIOCHEMISTRY 24:4512-4520(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89214217.
RA   WINES D.R., BRADY J.M., PRITCHETT D.B., ROBERTS J.L., MACDONALD R.J.;
RT   "Organization and expression of the rat kallikrein gene family.";
RL   J. BIOL. CHEM. 264:7653-7662(1989).
RN   [3]
RP   SEQUENCE FROM N.A.
```

RT "Characterization of genes encoding rat tonin and a kallikrein-like
RT serine protease.";
RL BIOCHEMISTRY 28:5334-5343(1989).
RN [4]

RA MEDLINE; 87271223.
RA LAZURE C., LEDOC R., SEIDAH N.G., THIBAUT G., GENEST J., CHRETIEN M.
RT "The complete amino acid sequence of rat submaxillary gland tonin
RT .does contain the aspartic acid at the active site: confirmation by

RP BIOCHEM. CELL BIOL. 65:321-337(1987).
 RN [5]
 RP SEQUENCE OF 25-103 AND 120-259.
 RN
 RX MEDLINE: 84117504.
 RA LAZURE C., LEDIC R., SEIDAN N.G., THIBAUT G., GENEST J., CHRETIEN M.
 RT "amino acid sequence of rat submaxillary tonin reveals similarities
 RT to serine proteases.".
 RL NATURE 307:555-558(1984).
 RN [6]
 RP SEQUENCE OF 25-34.
 RX MEDLINE: 90147705.
 RA KAMADA M., FURUHATA N., YAMAGUCHI T., IKEYITA M., KIZUKI K.,
 RA MORITA H.;
 RT "observation of tissue prokallikrein activation by some serine
 RT proteases, arginine esterases in rat submandibular gland.".
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 166:231-237(1990).
 RN [7]
 RP SEQUENCE OF 25-50, AND CHARACTERIZATION.
 RX MEDLINE: 92250562.
 RA MOREAU T., BRILLAUD-BOURDET M., BOUCHNIK J., GAUTHIER F.;
 RT "protein products of the rat kallikrein gene family. Substrate
 RT specificities of kallikrein kK2 (tonin) and kallikrein kK9.".
 RL J. BIOL. CHEM. 267:10045-10051(1992).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).

RX MEDLINE: 88011252.
 RA FUJINAGA M., JAMES M.N.G.;
 RT "Rat submaxillary gland serine protease, tonin. Structure solution
 and refinement at 1.8-A resolution.";
 RL J. MOL. BIOL. 195;373-396(1987)
 CC -1- FUNCTION: THIS PROTEIN HAS BOTH TRYPSIN- AND CHYMOTRYPSIN-LIKE
 ACTIVITIES, BEING ABLE TO RELEASE ANGIOTENSIN II FROM ANGIOTENSIN
 I OR ANGIOTENSINOGEN.
 CC -1- SUBUNIT: MONOMER.
 CC -1- TISSUE SPECIFICITY: FOUND IN SUBMAXILLARY GLAND.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M1565; G205034; -;
 DR EMBL: M23878; G207413; -;
 DR EMBL: M23877; G207413; JOINED.
 DR EMBL: M26533; G206776; -;
 DR PIR: A00945; KORTIN.
 DR PIR: A30971; A30971.
 DR PIR: A34050; A34050.
 DR PIR: C23863; C23863.
 DR PIR: B33359; B33359.
 DR PIR: A32340; A32340.
 DR PDB: 1TON; 16-JAN-88.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PRAM: PF00089; trypsin; 1.
 KW HYDROLASE; SERINE PROTEASE; GLYCOPROTEIN; MULTIGENE FAMILY; ZMOGEN;
 KW SIGNAL; 3D-STRUCTURE.
 FT SIGNAL 1 18
 FT PROPEP 19 24
 FT CHAIN 25 259
 FT ACT_SITE 63 63
 FT ACT_SITE 118 118
 FT ACT_SITE 211 211
 FT DISULFID 31 171
 FT DISULFID 48 64
 FT DISULFID 150 217
 FT DISULFID 182 196
 FT DISULFID 207 232
 FT CARBOHYD 106 106
 FT CARBOHYD 189 189
 FT STRAND 26 26
 FT STRAND 29 30
 FT TURN 33 34
 FT TURN 37 38
 FT STRAND 39 43
 FT STRAND 47 54
 FT TURN 55 56
 FT STRAND 57 60
 FT HELIX 62 64
 FT STRAND 70 73
 FT STRAND 77 77
 FT TURN 78 79
 FT TURN 83 84
 FT TURN 86 88
 FT STRAND 90 95
 FT TURN 97 98
 FT TURN 116 117
 FT STRAND 120 124
 FT TURN 146 147
 FT STRAND 149 154
 FT STRAND 168 168
 FT STRAND 170 177
 FT HELIX 179 181

FT HELIX 183 186
 FT TURN 188 189
 FT HELIX 190 193
 FT STRAND 194 198
 FT TURN 200 201
 FT STRAND 205 205
 FT TURN 208 209
 FT TURN 211 212
 FT STRAND 214 217
 FT TURN 218 219
 FT STRAND 220 225
 FT TURN 235 236
 FT STRAND 239 243
 FT HELIX 244 257
 SQ SEQUENCE 259 AA; 28248 MW; 745DA500 CRC32;

 alignment_scores:
 Quality: 246.50 Length: 161
 Ratio: 2.541 Gaps: 5
 Percent Similarity: 60.248 Percent Identity: 34.783

 alignment_block:
 US-09-030-606-177 x KLR2_RAT ..

 Align seg 1/1 to: KLR2_RAT from: 1 to: 259

 4 CACTCGCACCCCTGGCAGGGCGGACGTGTCATGGAACCATTTGTTCTG 53
 ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 34 AsnSerGlnProTyrGlnValAlaValIle.....AsnGluTyrLeuCy 48

 54 CTCGGCGCTCGTGGTCATCCGACAGTGGGTCGTGTCACCGCACAGTGT 103
 :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 48 scIglValIleuIleAspProSerTyrValIleThrAlaIleHisCysT 65

 104 TCAGAACCTCTACACCATCGGCGCTGGGCTGACAGTGTGAGCGCGAC 153
 :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 65 YrSerAsnAsnTyrGlnValIleuIleuGlyArgAsnAsnLeuPheIle 81

 154 CAAGAGCCAGGAGGAGGACAGATGTCGAGGCGCCTCTCCGACGCGAC 203
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 82 ..GluProPheAlaGlnArgGlyLeuValArgGlnSerPheArgHisP 97

 204 AGAGTACAAACAGACCTGTGTC..... 225
 :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 97 cAsPtyr...IleProIleuIleValIleHisAsnSerPheGluGlnProVal 113

 226GCTAACGACCTCATGCTCATCAAGTTGACGAATCCGTGTC 267
 :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 113 IsAspHisSerAsnAspLeuMetLeuHisLeuSerGluProAlaAsp 129

 268 GAGTGTGACACCATCCGAGCATGATGTCGTCGAGTCCCTACCGC 317
 :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 130 IleThrGlyGlyValIleValIleAspLeuProThrGlyGluProLysVa 146

 318 GGGGACCTCTGCTCGTTCTGCGTGGGT..... 348
 :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 146 lGlySerThrCysLeuAlaSerGlyTyrPglySerThrAsnProSerGlu 163

 349CTGCTGGCGAAC 360
 :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 163 etValIleSerHisAspLeuGlnCysValAsnIleHisLeuSerAsn 179

 361 GATGCTGATGTCACATCCACATGCCAGACTGTG 393
 :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 180 GluLysCysIleGluThrTyrLysAspAsnVal 190

 seq_name: SwissProt_37:KLR2_RAT
 seq_documentation_block:
 ID KLR2_RAT STANDARD; PRT; 261 AA.
 AC P36373;
 DT 01-JUN-1994 (REL. 29, CREATED)

OM of: US-09-030-606-177 to: SPTREMBL_10.* out_format: pfs
 Date: Sep 25, 1999 10:35 AM

About: Results were produced by the Gencore software, version 4.5,
 Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
 -MODEL=fram+np2.model -DEV=xlp
 -O=/gen2.1/USPTO.spool/US09030606/runat.24091999.171617_29869/app-query.fasta.1
 -DB=SPTREMBL_10 -OPMT=fastan -SUFFIX=rspt -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELDP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blissum62
 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=score
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
 -NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:
 Query: US-09-030-606-177
 Query length: 1119
 Database: SPTREMBL_10.*
 Database sequences: 201082
 Database length: 61543640
 Search time (sec): 260.540000

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
sp_rident:0920M1	469.50	825.38	1.7e-38	255	0920M1 mus musculus (mouse). en
sp_human:015096	275.00	478.86	3.8e-19	234	015096 homo sapiens (human). pr
sp_mammal:029474	271.50	471.64	8.6e-18	261	029474 canis familiaris (dog). pr
sp_human:015846	269.00	468.55	1.5e-18	223	015846 homo sapiens (human). pr
sp_human:075837	267.50	463.81	2.2e-18	282	075837 homo sapiens (human). se
sp_human:015098	265.00	463.48	3.7e-18	176	015098 homo sapiens (human). pr
sp_human:016272	265.00	462.85	3.7e-18	189	016272 homo sapiens (human). pr
sp_rident:063274	256.00	444.85	3.0e-17	235	063274 rattus norvegicus (rat). pr
sp_vertebrate:042158	255.50	441.75	4.2e-17	247	042158 petromyzon marinus (sea
sp_vertebrate:042608	253.50	439.85	5.3e-17	247	042608 petromyzon marinus (sea
sp_vertebrate:042160	248.50	431.08	1.7e-16	245	042160 petromyzon marinus (sea
sp_rident:003955	248.00	430.01	1.9e-16	250	003955 prionomys natalensis (afri
sp_rident:063275	247.50	429.51	2.1e-16	239	063275 rattus norvegicus (rat). pr
sp_rident:088309	245.00	424.27	3.8e-16	261	088309 mus musculus (mouse). pr
sp_rident:092189	245.00	421.21	5.9e-16	246	092189 mus musculus (mouse). pr
sp_vertebrate:042159	242.50	420.39	6.6e-16	244	042159 petromyzon marinus (sea
sp_human:015665	242.50	419.87	6.7e-16	259	015665 homo sapiens (human). tr
sp_vertebrate:093265	239.00	413.92	1.5e-15	250	093265 pleurocytes americanus
sp_rident:061855	235.00	406.39	3.7e-15	261	061855 mus musculus (mouse). tr
sp_rident:054854	231.50	400.48	8.3e-15	251	054854 rattus norvegicus (rat). pr
sp_rident:088301	228.00	394.40	1.8e-14	246	088301 mus musculus (mouse). se
sp_vertebrate:092099	227.00	392.75	2.3e-14	242	092099 paramecium magellani
sp_vertebrate:093266	223.00	385.60	5.8e-14	242	093266 pleurocytes americanus
sp_vertebrate:092046	223.00	385.35	5.8e-14	249	092046 dissonichus mawsoni. pr
sp_vertebrate:092046	222.50	384.89	6.5e-14	237	092046 dissonichus mawsoni. pr
sp_vertebrate:093594	212.50	369.52	6.2e-13	178	093594 dicentrarchus labrax (eu
sp_human:099885	209.50	360.57	1.3e-12	268	099885 homo sapiens (human). ca
sp_vertebrate:091036	207.00	357.87	2.2e-12	219	091036 gadus morhua (atlantic c
sp_rident:008643	205.50	353.96	3.2e-12	264	008643 mus musculus (mouse). pr
sp_human:000765	205.50	353.96	3.2e-12	252	000765 homo sapiens (human). ch
sp_vertebrate:011059	201.50	346.60	8.1e-12	258	011059 trimerusus graminus (ch
sp_mammal:046683	201.00	345.95	1.0e-12	251	046683 ovis aries (sheep). mat
sp_rident:063637	200.50	345.41	1.0e-11	241	063637 rattus norvegicus (rat). pr
sp_rident:061388	199.50	343.23	1.3e-11	252	061388 mus musculus (mouse). cy
sp_vertebrate:091053	199.00	342.00	1.4e-11	262	091053 glycydus ussuriensis. ca
sp_vertebrate:092077	199.00	341.86	1.4e-11	266	092077 gadus morhua (atlantic c
sp_vertebrate:093632	195.00	334.91	3.6e-11	260	093632 glycydus halys (halys v
sp_vertebrate:034289	194.50	334.19	4.0e-11	255	034289 salvelinus fontinalis (ch
sp_human:016018	193.50	333.46	5.0e-11	226	016018 homo sapiens (human). ch
sp_vertebrate:0916S1	193.00	332.26	5.6e-11	234	0916S1 agkistrodon acutus (dein
sp_rident:063636	193.00	331.75	5.6e-11	248	063636 rattus norvegicus (rat). pr
sp_vertebrate:09Y616	193.00	331.34	5.7e-11	260	09Y616 agkistrodon halys pallasi

sp_rident:063224 + 192.00 329.96 7.1e-11 248 063224 rattus norvegicus (ra
 sp_human:014243 + 190.00 325.68 1.1e-10 269 014243 homo sapiens (human).
 sp_mammal:019023 + 189.50 325.18 1.3e-10 257 019023 macaca mulatta (rhesu

seq_name: sp_rident:0920M1

seq_documentation_block:
 ID 0920M1 PRELIMINARY; PRT; 255 AA.

AC 0920M1;
 DT 01-MAY-1999 (TREMBLER, 10, Created)
 DT 01-MAY-1999 (TREMBLER, 10, Last sequence update)
 DE ENAMEL MATRIX SERINE PROTEINASE 1 PRECURSOR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SWISS-WEBSTER;
 RA Simmer J.;
 RT "Enamel Matrix Serine Proteinase 1 (EMSP1).";
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL, AF019979; AAC98894.1; -;
 KW Signal.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 32 255 MATRIX SERINE PROTEINASE 1.
 SQ SEQUENCE 255 AA; 27488 MW; 6E711616 CRC32;

alignment_scores:
 Quality: 469.50 Length: 139
 Ratio: 4.083 Gaps: 1
 Percent Similarity: 82.734 Percent Identity: 63.309

alignment_block:
 US-09-030-606-177 x 0920M1 ..

Align seg 1/1 to: 0920M1 from: 1 to: 255

4 CACATCGACGCGCTGCGACGCGCGACGTGTCATGGAACGATGTTCTG 53
 |||||
 41 HISSERGINPOTRPGlnAlaAlaLeupheseRglunspglpPhepocy 57
 |||||
 54 CTCGGGCGTCCTGGTGCATCCGAGTGGGCTCTGACGCGCACAGTGT 103
 |||||
 57 SSGEGLYAlaLeuValHisProGlnTIPValLeuSerAlaAlaHisCysL 74
 |||||
 104 TCAGAACCTCTACACCATCGGCGTGGCGCTGCACAGTCTTGAGCGGAC 153
 |||||
 74 euglnGlnSerTyrIleValAlaGlyLeuGlnHisAsnLeuGlySer 90
 |||||
 154 CAAGAGCCAGGAGCCAGATGCTGAGGCGCCTCTCCGACGCGCACCC 203
 |||||
 91 GlnGlnProGlnSerArgMetLeuGlnAlaHisLeuSerIleGlnHisPr 107
 |||||
 204 AGAGACACAGACAGACCTGCTGCGTACAGACCTCATGCTCAAGTTGG 253
 |||||
 107 OAsnPhenAsnProSerPheAlaAsnAspLeuMetLeuIleLeuAla 124
 |||||
 254 AGGATTCGCTCGAGTGCATGCACACCATCCGAGCATCAGCATGCTTCG 303
 |||||
 124 snGlnSerValIleGlnSerAsnThrIleArgSerIleProValAlaLeu 140
 |||||
 304 CAGTCCCTACCGCGGGAACCTCTGCTCTGCTGCTGGGCTGCTGCT 353
 |||||
 141 GlnCysProThrProGlnAspThrCysLeuValSerGlyTrpGlyGlnLe 157
 |||||
 354 GCGGACCATGCTGATGATGATGCATCAGATCCAGACATGCGAGGCTGGG 403
 |||||
 157 ulysAsnGlnLysLeuProSerLeu.....LeuG 167
 |||||
 404 AGGTGAGAGACTTTC 420
 |||||

Percent Similarity: 73.109 Percent Identity: 43.697

alignment_block:

US-09-030-606-177 x 075837 ..

Align seg 1/1 to: 075837 from: 1 to: 282

```

4 CACTCGCAGCCCTGGCAGGCGGCGCATGTGATAAGAAAGAAATTTCTG 53
|||||.....
63 HisSerLInProtrpGlnAlaAlaLeuPheGlnLysThrArgLeuLys 79
54 CTGGGGGCTCTGTGCATCCGAGTGGTGTCTCAGCCGACACTGTT 103
|||||.....
79 SGlyAlaThrLeuLeuAlaProArgTrpLeuThrAlaAlaHisCysL 96
104 TCCAGAACTCTACACCATCGGGCTGGGCTGCAGCTTGTGAGCGGAC 153
|||||.....
96 euLysProArgTrpLLeuAlaHisLeuGlyGlnHisAsnLeuGln..Lys 111
154 CAAGAGCCAGGAGCCAGATGTGGAGCCAGCCCTCCGTACGGCACC 203
|||||.....
112 GlnGlnLysSerGlnGlnThrArgThrAlaThrGlnSerPheProHisPr 128
204 AGAGTACAAAGACCCCTG.....CTCGTAAGACCTCATGC 241
|||||.....
128 OGlyPheAsnAsnSerLeuProAsnLysAspHisArgAsnAsnLysMetL 145
242 TCATCAAGTGGAGCAATCCGTGCGAGTCTGACACCATCCGAGGATC 291
|||||.....
145 euValLysMetAlaSerProValSerLLeuThrTrpAlaValArgProLeu 161
292 AGCATTCCTTCGACAGTCCCTACCGCGGGAAGCTTGCCTGTTCTGG 341
|||||.....
162 ThrLeuSerSerArgCysValThrAlaGlyThrSerCysLeuIleSerG 178
342 CTGGGGT 348
|||||
178 YTrpGly 180

```

seq_name: sp_human:Q15098

seq_documentation_block:

```

ID Q15098 PRELIMINARY: PRT: 176 AA.
AC Q15098;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-NOV-1996 (TREMBLrel. 08, Last annotation update)
DE PROSTATE SPECIFIC ANTIGEN (PSA) PRECURSOR (FRAGMENT).
GN APS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88326297.
RA RIESMAN P.H., KLAASSEN P., DER KOPPUT J.A., ROMIJN J.C., TRAPMAN J.;
RT "Molecular cloning and characterization of novel prostate antigen
RT cDNA's.";
RL Biochem. Biophys. Res. Commun. 155:181-188(1988).
DR EMBL: M21897; AAA5997.1; -.
DR PFM: PF00089; trypsin.1.
FT NON_TER 1
FT SIGNAL <1 13 POTENTIAL.
FT CHAIN 21 176 POTENTIAL.
SQ SEQUENCE 176 AA; 19158 MW; 949BC6C3 CRC32;

```

alignment_scores:

Quality: 265.00 Length: 126
Ratio: 3.011 Gaps: 2
Percent Similarity: 69.841 Percent Identity: 42.063

alignment_block:

US-09-030-606-177 x Q15098 ..

Align seg 1/1 to: Q15098 from: 1 to: 176

```

4 CACTCGCAGCCCTGGCAGGCGGCGCATGTGATAAGAAAGAAATTTCTG 53
|||||.....
30 HisSerLInProtrpGlnAlaAlaLeuPheGlnLysThrArgLeuLys 46
54 CTGGGGGCTCTGTGCATCCGAGTGGTGTCTCAGCCGACACTGTT 103
|||||.....
46 SGlyGlyValLeuValHisProGlnTrpValLeuThrAlaAlaHisCysL 63
104 TCCAGAACTCTACACCATCGGGCTGGGCTGCAGCTTGTGAGCGGAC 153
|||||.....
63 LeuArgAsnLysSerValIleLeuLeuGlyArgHisSerLeuPheHisPro 79
154 CAAGAGCCAGGAGCCAGATGTGGAGCCAGCCCTCCGTACGGCACC 203
|||||.....
80 GlnAspThrGly...GlnValPheGlnValSerHisSerPheProHisPr 95
204 AGAGTACAAAGACCCCTGCTCGCTAAC.....
95 OLeuTrpAspMetSerLeuLeuLysAsnArgPheLeuArgProGlyAspA 112
232 .....GACCTCATGCTCATCAAGTGGAGCAATCCGTGTCGAG 270
|||||.....
112 spSerSerHisAspLeuMetLeuValArgLeuSerGlnProAlaGlnLeu 128
271 TCAGACACCATCCGAGCATGACATGCTTGCATGCTGCCCTACCGCGG 320
|||||.....
129 ThrAspAlaValLysValMetAspLeuProThrGlnGlnProAlaLeuG 145
321 GAACCTTCGCTGCTTCTGCGTGGGT 348
|||||.....
145 YThrThrCysTrpAlaSerGlyTrpGly 154

```

seq_name: sp_human:Q16272

seq_documentation_block:

```

ID Q16272 PRELIMINARY: PRT: 189 AA.
AC Q16272;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE PSA PROTEIN (FRAGMENT).
GN PSA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95079406.
RA MONNE M., CROCE C.M., YU H., DIAMANDIS E.P.;
RT "Molecular characterization of prostate-specific antigen messenger
RT RNA expressed in breast tumors.";
RL Cancer Res. 54:6344-6347(1994).
DR EMBL: S75755; A014185.1; -.
DR PFM: PF00089; trypsin.1.
FT NON_TER 1
SQ SEQUENCE 189 AA; 21039 MW; DB4C6482 CRC32;

```

alignment_scores:

Quality: 265.00 Length: 126
Ratio: 3.011 Gaps: 2
Percent Similarity: 69.841 Percent Identity: 42.063

alignment_block:

US-09-030-606-177 x Q16272 ..
Align seg 1/1 to: Q16272 from: 1 to: 189

4 CACTCGCAGCCCTGGCAGGCGGCGCATGTGATAAGAAAGAAATTTCTG 53

```

|||||.....
47 HisserginProtrpglnValleuValalaserarglyArgalalValcy 63
54 CTGGGGCTCTGGTGCATCCGAGTGGTGTCTGCACCCGACACTGT 103
63 sglvcllyallleuValhlsProglntprvalleuthralalahlscyst 80
104 TCGCAACTCTACACATCGGGCTGGGCTGCAGTCTTGAGCGGCAC 153
80 leargAsllysservallleuLeuEnglyArgHlsSerleupherhlsPro 96
154 CAAGACCCAGAGAGCCAGATGGAGCGACCTCTCCGATCGGACCC 203
97 Glusprlrgly...GlnValrphelglnValSerHlsSerPheProHlsPr 112
204 AGAGTACACAGACCTTGTCTGGCTGAC..... 231
112 OleulrYrsmetSerleuLeuLysAsnArgPheLeuArgProglYAspA 129
232 .....GACCTCATGCTCATCAAGTTGAGCAATCCGTCCGAG 270
129 spserSerHlsAspleuMetleuLeuArgleuSerGlnProalaglueu 145
271 TCTGACACCATCCGAGCATGAGCATGCTTCGAGTCCCTACCGCGGG 320
146 ThrAspAlaVallyValMetAspLeuProThrGlnGlnProAlaLeuGl 162
321 GACCTTGCTCTGTCTGTGGTGGGT 348
162 YThrThrcysTyralaserGlyTrpGly 171
seq_name: sp_rudent:063274

```

```

seq_documentation_block:
ID 063274 PRELIMINARY; PRT: 235 AA.
AC 063274;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE KALLIKREIN (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-KIDNEY;
RA ZINTZ C.B., MA J.X., CHAO J., CHAO L.;
  Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: L33839; AAA58781.1; -.
DR PIRAM: PF00089; trypsin; 1.
FT NON_TER 1
SQ SEQUENCE 235 AA; 26226 MW; E6DE5AD8 CRC32;

```

```

alignment_scores:
Quality: 256.00 Length: 126
Ratio: 2.943 Gaps: 3
Percent Similarity: 69.048 Percent Identity: 41.270

```

```

alignment_block:
us-09-030-606-177 x 063274 ..

```

```

Align seg 1/1 to: 063274 from: 1 to: 235
4 CACTCGACCCCTGGCAGCGGCGACTGTCATGAAAGCAATTGTCTG 53
10 AsnserGlnProtrpglnValalValale.....AsnArgTylency 24
54 CTGGGGCTCTGGTGCATCCGAGTGGTGTCTGCACCCGACACTGT 103
24 sglvcllyallleuLeuAspProserTrpValleThrAlaAlaHlsCysT 41
104 TCGCAACTCTACACATCGGGCTGGGCTGCAGTCTTGAGCGGCAC 153

```

```

..... ||| :: ||||| |||..... |||
41 YrSerHlsTyTrYrHlsValleuLeuEnglyArgHlsAsnLeuPhegluAsp 57
154 CAAGACCCAGAGAGCCAGATGGTGGAGCCGACCTCTCCGATCGGACCC 203
58 ...GlnProPheAlaGlnTyArgPheValSerGlnSerPheProHlsPr 73
204 AGAGTACACAGACCTTGTCTGTC..... 225
73 AsprTyAsnProPheLeuMetArgAsnHlsThrArgGlnThrclYrYrA 90
226 .....GCTACGACCTCATGCTCATCAAGTTGAGCAATCCGTCCGAG 270
90 sPrYrSerAsnAspleuMetleuLeuHlsLeuSerGlnProAlaAspIle 106
271 TCTGACACCATCCGAGCATGAGCATGCTTCGAGTCCCTACCGGGG 320
107 ThrAspGlyVallyValleAspLeuProThrGlnGlnProLysValGl 123
321 GACCTTGCTCTGTCTGTGGTGGGT 348
123 YSerThrcysLeuValaserGlyTrpGly 132
seq_name: sp_vertebrate:042158

```

```

seq_documentation_block:
ID 042158 PRELIMINARY; PRT: 247 AA.
AC 042158;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE TRYPSINOGEN A2 PRECURSOR.
GN TRYPA2.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Cephalaspidomorphi;
  Petromyzontiformes; Petromyzontidae; Petromyzon.
RN [1]
RP SEQUENCE FROM N.A.
RA ROACH J.C.;
  Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF011898; AAB69654.1; -.
DR PIRAM: PF00089; trypsin; 1.
KW Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 247 TRYPSIN A2.
SQ SEQUENCE 247 AA; 26309 MW; D049AA4C CRC32;

```

```

alignment_scores:
Quality: 254.50 Length: 143
Ratio: 2.447 Gaps: 4
Percent Similarity: 72.727 Percent Identity: 34.965

```

```

alignment_block:
us-09-030-606-177 x 042158 ..

```

```

Align seg 1/1 to: 042158 from: 1 to: 247
1 GCGCACTCGACCCCTGGCAGCGGCGACTGTCATGAAAGCAATTGT 50
32 AlaHlsSerGlnProtrpglnValalSerleuAsnHlsGlyTrHls...Ph 47
51 CTGGGGCTCTGGTGCATCCGAGTGGTGTCTGCACCCGACACT 100
47 ecyselYglYserleuLeuAsnSerGlnTrpValalSerAlaAlaHlsC 64
101 GTTCACAACTCC.....TACACATCGGGCTGGGCTGCAGAGTCT 144
145 GAGCGCCAGCAAGCCAGGAGCCAGATGGTGGAGGCCACTCTCGGT 194
81 PheValAsn...GlnGlyThrGlnGlnGlnHlsSerGlnAlaAla 96

```

```

195 ACGGACCCAGAGTACACAGACCCCTGCTCGCTAACGACCTCATGCTCA 244
    eginhlspglnltyrnserttrphtlreaspanaspllemtleui 113
245 TCAAGTTGACAGATCCGTCGAGTCTGACACATCCGAGCATGAC 294
    leylsleuserserprolathtleuansnlnlyralaglnalalea 129
295 ATGCTCGGAGTCCCTACCGGGGGAAGTCTGCTGCTTCTGCTG 344
    leuprosersercysvalasnthglyvalmetcysthrilleser 146
345 GGGT.....CTGCTGGCAACGATGCTGATTGCCATCC 379
    polylgluthtglnthrsertleeglserrproasvalleumec 163
146 polylgluthtglnthrsertleeglserrproasvalleumec 163
380 AGTCCAGACTGTGGAGGCTGGAGTGT 408
163 lnAlaProValleuserAspThrSerCys 172

```

seq_name: sp-vertebrate:042608

```

seq_documentation_block:
ID 042608 PRELIMINARY; PRT; 247 AA.
AC 042608;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE TRYPSINOGEN A3 PRECURSOR.
GN TRYP2.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Cephalaspidomorphi;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
RN [1]
RP SEQUENCE FROM N.A.
RA ROACH J.C.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RC TISSUE=ANTERIOR INTESTINE;
RA Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF011899; AAB69655.1; -
DR EMBL; AF011352; AAB65411.1; -
DR PFM; PF00089; trypsin; 1.
KW SIGNAL.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 247 TRYPSIN A3.
SQ SEQUENCE 247 AA; 26295 MW; 6D71AC2E CRC32;

```

alignment_scores: Quality: 253.50 Length: 143
Ratio: 2.438 Gaps: 4
Percent Similarity: 72.727 Percent Identity: 34.965

alignment_block:

US-09-030-606-177 x 042608 ..

Align seg 1/1 to: 042608 from: 1 to: 247

```

1 GCGGACTGCGAGCCCTGGAGGCGGCACTGTATGGAAGAAATGTT 50
    |||||||.....|
32 Alahlserrlnprotprglnvalserleuansnlnlelytyrhis...Ph 47
51 CTGCTCGGGGCTCTGTCATCCGAGTGGTGTCTGACGCCGACACT 100
    |||||||.....|
47 ecysglglyserleuileasergerlntrpvalserlalaahisc 64
101 GTTTCAGAACTCC.....TACACATCGGGGCTGGGCTGCACAGTCTT 144
    |||||||.....|
64 ystyrglnthralaserarglleserValarglleglylnhisasnile 80

```

```

145 GAGGCGACCAAGAGCCAGGAGCCAGATGTTGGAGGCCAGGCTTCGCT 194
    phevAlasn...glulYthrgluglnlnleeglnAlaserlyslAla 96
195 ACGGACCCAGAGTACACAGACCCCTGCTCGCTAACGACCTCATGCTCA 244
    eginhlspglnltyrnserttrphtlreaspanaspllemtleui 113
245 TCAAGTTGACAGATCCGTCGAGTCTGACACATCCGAGCATGAC 294
    leylsleuserserprolathtleuansnlnlyralaglnalalea 129
295 ATGCTCGGAGTCCCTACCGGGGGAAGTCTGCTGCTTCTGCTG 344
    leuprosersercysvalasnthglyvalmetcysthrilleser 146
345 GGGT.....CTGCTGGCAACGATGCTGATTGCCATCC 379
    polylgluthtglnthrsertleeglserrproasvalleumec 163
146 polylgluthtglnthrsertleeglserrproasvalleumec 163
380 AGTCCAGACTGTGGAGGCTGGAGTGT 408
163 lnAlaProValleuserAspThrSerCys 172

```

seq_name: sp-vertebrate:042160

```

seq_documentation_block:
ID 042160 PRELIMINARY; PRT; 245 AA.
AC 042160;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE TRYPSINOGEN B2 PRECURSOR (FRAGMENT).
GN TRYPB2.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Cephalaspidomorphi;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
RN [1]
RP SEQUENCE FROM N.A.
RA ROACH J.C.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF011901; AAB69657.1; -
DR PFM; PF00089; trypsin; 1.
KW SIGNAL.
FT SIGNAL 1 1 POTENTIAL.
FT CHAIN 14 245 TRYPSIN B2.
SQ SEQUENCE 245 AA; 26001 MW; 071872F0 CRC32;

```

alignment_scores: Quality: 248.50 Length: 143
Ratio: 2.510 Gaps: 4
Percent Similarity: 69.231 Percent Identity: 34.965

alignment_block:

US-09-030-606-177 x 042160 ..

Align seg 1/1 to: 042160 from: 1 to: 245

```

1 GCGGACTGCGAGCCCTGGAGGCGGCACTGTATGGAAGAAATGTT 50
    |||||||.....|
30 Alahlserrlnprotprglnvalserleuansnlnlelytyrhis...Ph 45
51 CTGCTCGGGGCTCTGTCATCCGAGTGGTGTCTGACGCCGACACT 100
    |||||||.....|
45 ecysglglyserleuileasergerlntrpvalserlalaahisc 62
101 GTTTCAGAACTCC.....TACACATCGGGGCTGGGCTGCACAGTCTT 144
    |||||||.....|
62 ystyrglnthralaserarglleserValarglleglylnhisasnile 78

```

```

      64 yrrthrttytTgGlnValntrPreuulLylysAsnAnLeuys.. Glu 79
      154 CAGAGCCAGGAGGACCCAGATGTTGGTGGAGCCAGCCTTCCTACGGCACCC 203
      180 TtYgluPProserAlaGlnPheArgGluValSerLysAlaIleProHisPr 96
      204 AGAGTACAC.....AGACCT 220
      96 OG1PheAsnHeserAsnMetMetAspGlntrThrGlntrLysLysA 113
      221 TGCTGGCTAAAGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAG 270
      113 sPTYSerAsnAspLeuMetLeuLeuArgLeuLeuSerGluProAlaGluIle 129
      271 TCTGACCAACATCCGAGCAGCATGAGCATTCCTTCGAGTGCCTACCGGG 320
      130 ThrAspValValLysProIleAspLeuProThrGluIuProthrValGI 146
      321 GACTCTTGCCCTCGTTCTGCTGGCTGGGT 348
      146 ySerArgCysLeuAlaSerGlyTyrGly 155
seq_name: sp_rodent.063275
seq_documentation_block:
ID 063275 PRELIMINARY; PRT; 239 AA.
AC 063275;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE KALIKREIN (FRAGMENT).
OS Rattus norvegicus (Rat.).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
NC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-SUBMANDIBULAR GLAND;
RA ZI2T2 C.B., MA J.X., CHAO J., CHAO L.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: L33840; AAA5782.1; -.
PFAM: PF00089; trypsin; 1.
FT NON_TER 1 1
SQ SEQUENCE 239 AA; 26382 MW; 0609E3B6 CRC32;

alignment_scores:
      Quality: 247.50      Length: 160
      Ratio: 2.500      Gaps: 6
      Percent Similarity: 61.875      Percent Identity: 35.000

alignment_block:
US-09-030-606-177 x 063275 ..

Align seg 1/1 to: 063275 from: 1 to: 239

      4 CACTCGACCCCTGGCAGGCGGCACTGGTCATGTGMAAAGCAATTGTCG 53
      18 AsnSerGlnProTyrGlnValAlaValIle.....AsnGluAspLeuGly 32
      54 CTCGGGGCTCTGCTGTGCATCCGCAAGTGGGTGCTGTACAGCGCACACTGTT 103
      32 sGlyIleValLeuIleAspProSerTrpValIleThrAlaAlaHisCysT 49
      104 TCCCAACATCCGTACACCATCGGTGGGCGCCGACAGACTTGTGAGGCGAC 153
      49 yTSeRAspAsnTyrHisValLeuLeuGlyGlnAsnAsnLeuSerGluAsp 65
      154 CAGAGCCAGGAGGAGGACAGATGTTGGTGGAGCGACCTTCCTGAGCGCACCC 203
      66 ValGln.....HisArgLeuVal.....SerIleSerPheArgHisPr 78
      204 AGAGTACACAGACCCCTGCTC..... 225

```

```

      78 OASPTY...LysProPheLeuMetArgAsnHisThrArgLysProLysA 94
      226 .....GCTAACGACCTCATGCTCATCAAGTTGGAGCAATCCGTGCGAG 270
      94 sPTYSerAsnAspLeuMetLeuHisLeuSerGluProAlaAspIle 110
      271 TCTGACACCATCGGAGCATCATGCTGCTGCTGAGTCCGCTACCGCGGG 320
      111 ThrAspGlyValLysValIleAspNeuroThryLysGluProLysValI 127
      321 GAACCTTGCCCTGCTTCTGCTGGGGT.....CTGCTGGCGACGAT 363
      127 ySerThrCysLeuValSerGlyTyrGlySerThrAsnProSerGluTyrG 144
      349 .....CTGCTGGCGACGAT 363
      144 IupheProAspAspLeuGlnCysValAsnIleHisLeuSerAsnGlu 160
      364 GCTGTGATTGCCATCCAGTCCGACGACTGTG 393
      161 LysCysIleLysAlaTyrLysGluLysVal 170

seq_name: sp_rudent:088309

seq_documentation_block:
ID 088309 PRELIMINARY; PRT; 261 AA.
AC 088309;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE PRORENIN-CONVERTING ENZYME (MK13B) PRECURSOR (MK13B).
GN MKR-13B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98351995.
RA HOSOI K., TRADA J., TSUMURA K., KANAMORI N., YAMANAKA N.;
RT "Expression of an allotype of prorenin-converting enzyme in the
RT submandibular gland of DBA/2N mice.";
RL J. Biochem. 124:368-376(1998).
DR EMBL: AB016032; BAA31686.1; -.
DR PIR: P00089; trypsin.1.
KW Signal.
FT SIGNAL 1 24 POTENTIAL.
SQ SEQUENCE 261 AA; 28692 MW; 8DB1814F CRC32;

```

alignment_scores:

Quality:	245.00	Length:	126
Ratio:	2.784	Gaps:	2
Percent Similarity:	69.841	Percent Identity:	38.889

alignment_block:

US-09-030-606-177 x 088309 ..

Align seg 1/1 to: 088309 from: 1 to: 261

```

4 CACTCGACCCCTGGCAGGCGGACACTGTCATGAAAGCAATTGTTCTG 53
.....|.....|.....|.....|.....|.....|.....|.....|
34 AsnSerGlnProTyrGlnValAlaValLysGlnLysGlnHisIleCys 50
54 CTCGGGCGCTGCTGTCATCCGCACTGGGTGCTGACGCCGACACTGTT 103
.....|.....|.....|.....|.....|.....|.....|.....|
50 sGlyGlyValLeuLeuAspArgAsnTyrValLeuThrAlaAlaHisCysT 67
104 TCACAAATCTCTACACCATCGGCGGCTGCGCTGACAGCTTTAGGCCGAC 153
.....|.....|.....|.....|.....|.....|.....|.....|
67 yTValAspGlnTyrGlnValLysLeuGlyLysAsnLysLeu...PheGln 82
..154-GATGAGCCAGGAGCCAGATGCTGAGAGCCAGCTCTCGCTAGCGACCC 203

```

```

      83 GluGluProSerAlaGlnHisArgLeuValSerLysSerPheProHispr 99
      204 AGAGTACACACGACCTTGCTC.....|.....|.....|.....|
      99 OGlyTyrAsnMetSerLeuLeuMetLeuGlnThrIleProProGlyAla 116
      226 .....GCTAACGACCTCATGCTCATCAAGTTGGAGCAATCCGTGCGAG 270
      116 sPheSerAsnAspLeuMetLeuLeuArgLeuSerLysProAlaAspIle 132
      271 TCTGACACCATCGGAGCATCATGCTGCTGCTGAGTCCGCTACCGCGGG 320
      133 ThrAspValLysProIleAlaLeuProThrLysGluProLysProGlu 149
      321 GAACCTTGCCCTGCTTCTGCTGGGGT 348
      149 ySerLysCysLeuAlaSerGlyTyrGly 158

seq_name: sp_rudent:Q921R9

seq_documentation_block:
ID Q921R9 PRELIMINARY; PRT; 246 AA.
AC Q921R9;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE TRYPSINOGEN 16.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRATN-BALB/C;
RA ROWEN L., HOOD L.;
RT "Comparison between strains Balb/C and 129 in a region of the mouse T
RT cell receptor beta locus.";
RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF107342; AAC79093.1; -.
SQ SEQUENCE 246 AA; 26134 MW; 36625744 CRC32;

```

alignment_scores:

Quality:	243.00	Length:	115
Ratio: <td>2.793</td> <td>Gaps: <td>2</td> </td>	2.793	Gaps: <td>2</td>	2
Percent Similarity: <td>75.652</td> <td>Percent Identity: <td>39.130</td> </td>	75.652	Percent Identity: <td>39.130</td>	39.130

alignment_block:

US-09-030-606-177 x Q921R9 ..

Align seg 1/1 to: Q921R9 from: 1 to: 246

```

4 CACTCGACCCCTGGCAGGCGGACACTGTCATGAAAGCAATTGTTCTG 53
.....|.....|.....|.....|.....|.....|.....|.....|
33 AsnSerValProTyrGlnValSerLeu...AsnSerGlyTyrHisPheCy 48
54 CTCGGGCGCTGCTGTCATCCGCACTGGGTGCTGACGCCGACACTGTT 103
.....|.....|.....|.....|.....|.....|.....|.....|
48 sGlyGlySerLeuLeuAsnAspGlnTyrValValSerAlaAlaHisCysT 65
104 TCACAAATCTCTACACCATCGGCGGCTGCGCTGACAGCTTTAGGCCGAC 153
.....|.....|.....|.....|.....|.....|.....|.....|
65 yTyrThrArgIleGlnValAlaValLeuGlnLysAsnIleAsnVal... 80
154 CAAGAGCCAGGAGCCAGATGCTGAGAGCCAGCTCTCGCTAGCGACCC 203
.....|.....|.....|.....|.....|.....|.....|.....|
81 LeuGlnGlyAsnGlnGlnPheIleAspAlaAlaLysIleIleLysHispr 97
204 AGAGTACACGACCTTGCTGCTGCTGAGTACGACCTCATGCTCAAGTTGG 253
.....|.....|.....|.....|.....|.....|.....|.....|
97 OAsnPheAsnArgLysThrLeuAsnAsnAspIleMetLeuLeuLysLeus 114
..254-ACGAATCCGTCTCGAGTCTGACACCATCGGAGCATCAGCATCAGCATGCTTGG 303

```

```
114 erSerProValThrLeuAsnAlaArgValAlaThrValAlaLeuProSer 130
304 CAGTGCCTACCGCGGAGACTCTGCCTCCTTCTGCTGGGT 348
131 SerCysAlaProAlaGlyThrGlnCysLeuSerGlyTTPGly 145
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:29:46 ; Search time 1809.22 seconds
(without alignments)
1220.009 Million cell updates/sec

Title: US-09-030-606-177
Sequence: 1119
1 GCCGACCTGCGAGCCCTGCA.....ATGTTAAAAA 1119

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST:*

1: em_est1:*

2: em_est2:*

3: em_est3:*

4: em_est4:*

5: em_est5:*

6: em_est6:*

7: em_est7:*

8: em_est8:*

9: em_est9:*

10: em_est10:*

11: em_est11:*

12: em_est12:*

13: em_est13:*

14: em_est14:*

15: em_est15:*

16: em_est16:*

17: em_est17:*

18: em_est18:*

19: em_est19:*

20: gb_est1:*

21: gb_est2:*

22: gb_est3:*

23: gb_est4:*

24: gb_est5:*

25: gb_est6:*

26: gb_est7:*

27: gb_est8:*

28: gb_est9:*

29: gb_est10:*

30: gb_est11:*

31: gb_est12:*

32: gb_est13:*

33: gb_est14:*

34: gb_est15:*

35: gb_est16:*

36: gb_est17:*

37: gb_est18:*

38: gb_est19:*

39: gb_est20:*

40: gb_est21:*

41: gb_est22:*

42: gb_est23:*

43: gb_est24:*

44: gb_est25:*

45: gb_est26:*

46: gb_est27:*

47: gb_est28:*

48: gb_est29:*

49: gb_est30:*

50: gb_est31:*

51: gb_est32:*

52: em_est20:*

53: em_est21:*

54: em_est22:*

55: em_est23:*

56: em_est24:*

57: em_est25:*

58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	361	32.3	777	48	AI557281	AI557281 PT2.1-15-
2	223.6	20.0	722	48	AI557025	AI557025 PT2.1-10-
3	214.2	19.1	415	35	AA551449	AA551449 nj55e05.s
4	205.8	18.4	569	50	AI686689	AI686689 tu35g11.x
5	170	15.2	216	32	AA336074	AA336074 ESTA0886
6	99.6	8.9	479	46	AI436093	AI436093 th92b12.x
7	99.6	8.9	475	46	AI432907	AI432907 t446g04.x
8	92	8.2	245	31	AA295609	AA295609 EST100780
9	91.6	8.2	267	32	AA366980	AA366980 EST77998
10	82.2	7.3	429	35	AA594076	AA594076 nm31a12.s
11	81.4	7.3	407	38	AA776018	AA776018 aeb3g11.s
12	81	7.2	523	35	AA542994	AA542994 n155b01.s
13	76.2	6.8	486	35	AA535837	AA535837 nj79g05.s
14	76.2	6.8	572	35	AA579026	AA579026 nf34f02.s
15	74.8	6.7	253	20	T29510	T29510 EST82448 Hu
16	74	6.6	546	36	AA388907	AA388907 mp16b03.r
17	72.6	6.5	336	34	AA506459	AA506459 nh46e08.s
18	72.6	6.5	548	35	AA578976	AA578976 nt26g03.s
19	72.6	6.5	316	35	AA593245	AA593245 nm07e10.s
20	72.6	6.5	902	47	AI525832	AI525832 PT1.3-06-
21	72.6	6.5	732	48	AI547309	AI547309 PN001_AH-
22	71.6	6.4	470	47	AI525128	AI525128 promena-7
23	71	6.3	610	48	AI557591	AI557591 pt2.1-1.F
24	70	6.3	371	34	AA503943	AA503943 nh38d03.s
25	70	6.3	484	36	AA603529	AA603529 np15b08.s
26	70	6.3	523	37	AA639901	AA639901 np08f03.s
27	70	6.3	737	48	AI547285	AI547285 PN001_AH-
28	70	6.3	454	50	AI685510	AI685510 tu36b10.x
29	69.8	6.2	703	48	AI546857	AI546857 pn2.1-07-
30	69.4	6.2	457	45	AI185433	AI185433 m185b05.y
31	69.4	6.2	595	48	AI557389	AI557389 PT2.1-6-G
32	69	6.2	619	26	W58737	W58737 zd23e07.r1
33	69	6.2	435	29	AA177005	AA177005 nc03b02.s
34	68.8	6.1	432	38	AA778667	AA778667 af87e12.s
35	67.8	6.1	340	27	AA038537	AA038537 m185b05.r
36	67.2	6.0	519	42	AI098337	AI098337 v984b03.r
37	67.2	6.0	760	42	AI127299	AI127299 qb74f07.x
38	66.8	6.0	453	34	AA533652	AA533652 nj72g08.s
39	65.6	5.9	616	48	AI546898	AI546898 pn2.1-09-
40	65.4	5.8	457	42	AI078417	AI078417 cz05e02.x
41	65	5.8	624	47	AI525001	AI525001 promena-4
42	64.2	5.7	397	35	AA573604	AA573604 nf42d12.s
43	64	5.7	472	28	AA062294	AA062294 m170a06.r
44	64	5.7	482	28	AA066464	AA066464 ml73c10.r
45	64	5.7	568	28	AA068804	AA068804 mm62g08.r

ALIGNMENTS

RESULT 1

AI557281 777 bp mRNA EST 23-MAR-1999

LOCUS PT2.1-15

DEFINITION PT2.1-15_G12.r tumor2 Homo sapiens cDNA 3', mRNA sequence.

ACCESSION AI557281

NID 94489644

VERSION AI557281.1 GI:4489644

KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 777)
AUTHORS Huang, G.M., Ng, W., Farkas, J., Chen, L., Liang, H.A., Gordon, D., Jun Yu, J., and Hood, L.
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138767.

Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of
Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.

FEATURES
source Location/Qualifiers

1..777
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="960H11; 6; 6p21.31-6p22.1; 21q"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional
cDNA library was constructed using Lambda ZP II kit
(Stratagene). mRNA was extracted from a frozen prostate
tumor tissue (Mayo Clinics)."
BASE COUNT 162 a 205 c 227 g 157 t 26 others
ORIGIN

Query Match 32.3%; Score 361; DB 48; Length 777;

Best Local Similarity 100.0%; Pred. No. 6.9e-96;
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCACTCCGAGCCCTGCGAGCGGCACTGTCATGAAAACGAATGTTCTGCTCGGGC 60
DB 14 GCCCATCTCGAGCCCTGCGAGCGGCGCACTGTCATGAAAACGAATGTTCTGCTCGGGC 73
QY 61 GTCTGGTGCATCCGAGTGGTGTCTGTACGCCGACACTGTTCCAGAACTCCTACACC 120
DB 74 GTCTGGTGCATCCGAGTGGTGTCTGTACGCCGACACTGTTCCAGAACTCCTACACC 133
QY 121 ATGGGGTGGGCTGCGACACTGTTGAGGGCGGACCAAGAGCCAGGAGCCAGATGGTGGAG 180
DB 134 ATGGGGTGGGCTGCGACACTGTTGAGGGCGGACCAAGAGCCAGGAGCCAGATGGTGGAG 193
QY 181 GCCAGCCTCCGCTAGCGGACCCAGAGTACAGACCTTGCTGCTAGCAGACCTCATG 240
DB 194 GCCAGCCTCCGCTAGCGGACCCAGAGTACAGACCTTGCTGCTAGCAGACCTCATG 253
QY 241 CTCATCAAGTTGAGCAATCCGTGTCCGAGTGTGACACCAATCCGGAGCAATGATTGCT 300
DB 254 CTCATCAAGTTGAGCAATCCGTGTCCGAGTGTGACACCAATCCGGAGCAATGATTGCT 313
QY 301 TCGCAGTGGCCTACCGCGGGGGAAGCTTGGCTGTTCTGGGCTGGCTGGCGGCAAC 360
DB 314 TCGCAGTGGCCTACCGCGGGGGAAGCTTGGCTGTTCTGGGCTGGCTGGCGGCAAC 373
QY 361 G 361
DB 374 G 374

RESULT 2
LOCUS A1557025 722 bp mRNA EST 23-MAR-1999
DEFINITION PT2.1.10_F05.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION A1557025
NID 94489388
VERSION 1 GI:4489388

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 722)
AUTHORS Huang, G.M., Ng, W., Farkas, J., Chen, L., Liang, H.A., Gordon, D., Jun Yu, J., and Hood, L.
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138511.

Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of
Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.

FEATURES
source Location/Qualifiers

1..722
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="15; 21q"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional
cDNA library was constructed using Lambda ZP II kit
(Stratagene). mRNA was extracted from a frozen prostate
tumor tissue (Mayo Clinics)."
BASE COUNT 146 a 156 c 195 g 176 t 49 others
ORIGIN

Query Match 20.0%; Score 223.6; DB 48; Length 722;

Best Local Similarity 95.6%; Pred. No. 2.1e-55;
Matches 237; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 109 AACTCTACACCATGGGCTGGGCGTGCACAGTCTGAGCGCGGACCAAGAGCAGGAGAC 168
DB 115 AGCTTTACACCATGGGCTGGGCGTGCACAGTCTGAGCGCGGACCAAGAGCAGGAGAC 174
QY 169 CAGATGGTGGAGCCAGCCTCTCCGTACGCGACCCAGAGTACAGACCTTGCTCGCT 228
DB 175 CAGATGGTGGAGCCAGCCTCTCCGTACGCGACCCAGAGTACAGACCTTGCTCGCT 234
QY 229 AACGACCTATGCTCATCAAGTTGAGCAATCCGTGTCCGAGTGTGACACCAATCCGGAGC 288
DB 235 AACGACCTATGCTCATCAAGTTGAGCAATCCGTGTCCGAGTGTGACACCAATCCGGAGC 294
QY 289 ATGAGCATGCTTCGAGAGGCCCTACCGGGGGAAGTCTTGGC-TGCTTTCGGCTGGGG 347
DB 295 ATGAGCATGCTTCGAGAGGCCCTACCGGGGGAAGTCTTGGC-TGCTTTCGGCTGGGG 354
QY 348 TCTGCTGG 355
DB 355 TCTGCTGG 362

RESULT 3
LOCUS AA551449 415 bp mRNA EST 05-SEP-1997
DEFINITION n155e05.s1 NCI_CGAP_Prg Homo sapiens cDNA clone IMAGE:296416
similar to SW:K1RA_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR
;; mRNA sequence.
ACCESSION AA551449
NID 92321701
VERSION AA551449.1 GI:2321701
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 415)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 9, 1995 this sequence version replaced gi:802243.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuquai,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrip/image/image.html

Insert Length: 640 Std Error: 0.00
Seq primer: -40m13 fwd. RT from Amersham
High quality sequence stop: 412.

FEATURES

Source

Location/Qualifiers
1..415
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:996416"
/clone_lib="NCI-CGAP_P19"
/sex="male"
/tissue_type="normal prostatic epithelial cells"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pAMP10; mRNA made from
normal prostatic epithelial cells; cDNA made by oligo-dT
priming; Non-directionally cloned. Size selected on
agarose gel, average insert size 600 bp. Library made by
D. Kitzman, NIH."

BASE COUNT 78 a 125 c 129 g 83 t

ORIGIN

Query Match 19.1%; Score 214.2; DB 35; Length 415;
Best Local Similarity 98.6%; Pred. No. 1e-52;
Matches 216; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 143 TTGAGCGGACCAAGAGCGAGGAGCCAGATGCTGAGGCGCCCTCCGACGGCAC 202
|||||
DB 2 TCGAGCGGACCAAGAGCGAGGAGCGAGGCGCCCTCCGACGGCAC 61
|||||
OY 203 CAGAGTACAGAGACCTTGCTGCTTAAGACCTCATGCTCAAGTTGAGCAATCCG 262
|||||
DB 62 CAGAGTACAGAGACCTTGCTGCTTAAGACCTCATGCTCAAGTTGAGCAATCCG 121
|||||
OY 263 TGTCCGAGTGTACACATCCGAGCATTCGATGCTTCGAGTCCCTACCGGGGGA 322
|||||
DB 122 TGTCCGAGTGTACACATCCGAGCATTCGATGCTTCGAGTCCCTACCGGGGGA 181
|||||
OY 323 ACTCTTGCTGCTTCTGCTGGGCTTCGCTGCGCAAG 361
|||||
DB 182 ACTCTTGCTGCTTCTGCTGGGCTTCGCTGCGCAAG 220
|||||

RESULT 4
LOCUS A1686689 569 bp mRNA EST 27-MAY-1999
DEFINITION tu35g11.x1 NCI-CGAP_P128 Homo sapiens cDNA clone IMAGE:2253092 3',
similar to TR:Q92046 Q92046 PREPROTRYPIN PRECURSOR ;, mRNA
sequence.
ACCESSION A1686689
NID 94897983
VERSION A1686689.1 GI:4897983
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 569)
AUTHORS Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189584.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrip/image/image.html

Seq primer: -40UP from Glibco
High quality sequence stop: 444.

FEATURES

Source

Location/Qualifiers
1..569
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2253092"
/clone_lib="NCI-CGAP_P128"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI-CGAP_P122 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonoids
985608-986759, 1101192-1101959, and 1217928-1220615)."
Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 125 a 150 c 170 g 124 t

ORIGIN

Query Match 18.4%; Score 205.8; DB 50; Length 569;
Best Local Similarity 94.7%; Pred. No. 3.4e-50;
Matches 213; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 157 GAGCAGGAGACCAAGATGCTGAGGCGCCCTCCGACGGCACCCAGATACAGCA 216
|||||
DB 569 GAGCAGGAGACCAAGATGCTGAGGCGCCCTCCGACGGCACCCAGATACAGCA 510
|||||
OY 217 CCGTTCGCTGCTTAAGACCTCATGCTCAAGTTGAGCAATCCGAGTCTGAC 276
|||||
DB 509 CCGTTCGCTGCTTAAGACCTCATGCTCAAGTTGAGCAATCCGAGTCTGAC 450
|||||
OY 277 ACCATCCGAGCATTCGATGCTTCGAGTCCCTACCGGGGGAACCTTGGCTGTT 336
|||||
DB 449 ACCATCCGAGCATTCGATGCTTCGAGTCCCTACCGGGGGAACCTTGGCTGTT 390
|||||
OY 337 TGTGGCTGGGCTGCTGCGGAAGAGAGCTGATGTCATCCAG 381
|||||
DB 389 TGTGGCTGGGCTGCTGCGGAAGAGAGCTGATGTCATCCAG 345
|||||

RESULT 5
LOCUS AA336074 216 bp mRNA EST 21-APR-1997
DEFINITION EST40886 Endometrial tumor Homo sapiens cDNA 5' end similar to
similar to kallikrein family, mRNA sequence.
ACCESSION AA336074
NID 91988560

VERSION AA36074.1 GI:1988560
 EST.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE 1 (bases 1 to 216)
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS Adams,M.D., Kerlavage,A.R., Fleischman,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-ai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fitchman,J.L., Geoghegan,N.S., Glodde,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.F., Ferris,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M., and Venter,J.C.
 TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
 JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
 MEDLINE 96026280
 COMMENT On Apr 14, 1993 this sequence version replaced gi:692773.

CONTACT: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423

Email: arkerlav@tigr.org
 For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tcbl/hgi/hgi.html>)
 Seq primer: M13 Reverse

FEATURES
 Location/Qualifiers

1..216
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):138014"
 /db_xref="taxon:9606"
 /clone_lib="Endometrial tumor"
 /sex="female"
 /dev_stage="adult"
 /note="Organ: endometrium; Vector: pBluescript SK-";
 Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 45 a 65 c 63 g 38 t 5 others
 ORIGIN

Query Match 15.2%; Score 170; DB 32; Length 216;
 Best Local Similarity 97.1%; Pred. No. 8.1e-40;
 Matches 170; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CGCAGTGCAGCCTGGCAGCGGCGACATGTCATGAAAGCAATGTTGCTCGGGCG 61
 |||
 DB 42 CGCAGTGCAGCCTGGCAGCGGCGACATGTCATGAAAGCAATGTTGCTCGGGCG 101
 |||
 QY 62 TCCGAGGACATCGCAGGAGGCTGTGTCAGCCGACACTGTTCCAGAACCCACACCA 121
 |||
 DB 102 TCCGAGGACATCGCAGGAGGCTGTGTCAGCCGACACTGTTCCAGAACCCACACCA 161
 |||
 QY 122 TCGGGCTGGGCTCGACAGTCTTGAGGCGGACAGAGCCAGGAGGACAGATGTT 176
 |||
 DB 162 TCGGGCTGGGCTCGACAGTCTTGAGGCGGACAGAGCCAGGAGGACAGATGTT 216
 |||

RESULT 6

AI436093/ 479 bp mRNA EST 30-MAR-1999
 LOCUS tb92b12.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 DEFINITION IMAGE:2126111 3' similar to TR:002710 002710 GAG POLYPROTEIN.;
 mRNA sequence.

ACCESSION AI436093
 NID 94307772
 VERSION AI436093.1 GI:4307772
 KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 479)
 NCICGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 18, 1998 this sequence version replaced gi:3137463.

CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAG Consortium (info@image.lnl.gov) for further information.
 Insert Length: 677 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 444.

FEATURES
 Location/Qualifiers

1..479
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2126111"
 /clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following NHP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and clones: Soares NBHP pool 1: 309384-310919, 323208-325895 Soares NBHP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NBHP pool 1: 758280-760583, 772104-774407 Soares NBHP pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHP pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 143 a 85 c 127 g 124 t
 ORIGIN

Query Match 8.9%; Score 99.6; DB 46; Length 479;
 Best Local Similarity 62.5%; Pred. No. 6.4e-19;
 Matches 240; Conservative 0; Mismatches 134; Indels 10; Gaps 5;

QY 557 AGTCAGACTATCATGATTACTGTGAGCTGCTGTATTGTACTAACCATGCC--- 613
 |||
 DB 458 AGACAGACTATTCGATTCGCTTCCTGCTGTCATGACAGAGTACACCCACC 399
 |||
 QY 613 --GATGTTTGGGAATTTAGCGTCTGCTCAACCATCTTGGTATCCATTTCT 670
 |||
 DB 398 TTGCTTTTGGGTTGAGTGTCTGCTCAACCATCTTGGTATCCATTTCT 339
 |||
 QY 671 CACGATTTAGATTTCTCTTCAGTGTGATCCATCCAC-ATAATTCTACCCAGCA 729
 |||
 DB 338 CATGTTTAAATTTTGTGATGAGTGTGTTCCACTGTTAGTCTGATCATACC 279
 |||
 QY 730 GAGGTGAGGATCATATAGCTCTCAAGATGCTGACTCCCTCAAAATTCATTCT 789
 |||

[illegible]

Seq primer: -40ml3 fwd. ET from Amerisham
High quality sequence stop: 384.
Location/Qualifiers
1. 429

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1085470"
/clone_1lb="NCI CGAP Gas1"
/tissue_type="gastric tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: stomach; Vector: Bluescript SK-; Site: 1; EcoRI; Site: 2; XhoI; Cloned unidirectionally. Primer: Oligo dt. Pooled gastric tumors. 5' adaptor sequence: 5' GAATGGGCGAG 3' 3' adaptor sequence: 5' CTGACATTTTCTTTTCTTTTCTTTT 3' Average insert size: 1.0 kb."
BASE COUNT 114 a 100 c 78 g 137 t
ORIGIN

Query Match 7.3%; Score 82.2; DB 35; Length 429;
Best Local Similarity 62.0%; Pred. No. 8.3e-14; Mismatches 199; Conservative 0; Indels 9; Gaps 4;

532 ACTAGCCAGCAGCAGCAGTCTT-CCGAGTGCAGATCATGATCTGTTGACGTGC 590
109 ATTAGCCAAAGCCAGAGCTGCGACAATCAGACTGTCATGCTTCTTGGCAGCT 168
591 TGTCTATTGACTACTACCATGCGCATGTTT-----AGGTGAATTAAGCGTCTGGCCT 644
169 TGTCCATTAGTAGTACAGTACACCCCTTTCTTGTAGTGTGAGTGCCTCCACCTGGGCC 228
645 CAACCATCTGTGATCCAGTATCCCTCAGTATGATGATTCCTGCTGCTGCTGACGCC 704
229 CTGCGACCTCAGGATTCATTAATTTCCATTTATTAATTTTGTAGTGTGACGTGAG 288
705 ATTCCAC-ATAATTTCTGACCTACAGAGTGTAGGAGATCATATGCTTCAAGATGCT 763
289 GTTCCACTGTATACCTGCGATACAGAAAGCAATTAAGCCCTCTTCAAGATGCA 348
764 GGTACTCCCTCACAATTCATTTCTCTGTGTAG-TGAAGTGGCCCTGTGAGCC 822
349 GATATGCTTTCACAAATCTATTTCACAAAGATTAATGATTAAGTATGTCTTGTGACC 408
823 TCCAGGAGTGGGTGACAGT 843
409 TCCAGCTGGGTGAGTAGT 429

RESULT 11
AA776018/c 407 bp mRNA EST 05-FEB-1998
DEFINITION a683611.s1 Stragene schizo brain S11 Homo sapiens cDNA clone
IMAGE:970820 3', mRNA sequence.
ACCESSION AA776018
NID 9283352
VERSION AA776018.1 GI:2835352
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 407)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,
Kizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
Therling, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
TITLE WashU-NCI human EST Project
COMMENT On Jan 17, 1998 this sequence version replaced gi:2044645.
Contact: Wilson RK
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: polyt not found
Seq primer: -40ml3 fwd. ET from Amerisham.
Location/Qualifiers
1. 407

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:970820"
/clone_1lb="Stragene schizo brain S11"
/sex="male"
/tissue_type="schizophrenic brain S-11 frontal lobe"
/dev_stage="34 years old"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: Bluescript SK-; Site: 1; EcoRI; Library
constructed from S-11 frontal lobe, male, 34 years old,
50% caucasian, 50% Aleutian. Schizophrenic suicide.
Random primed into EcoRI site of ZAP II vector. Mass
excised. Custom library. Avg insert length 1.4kb.
Material obtained by Johnston N., Torrey, E.F., Yolken R.,
and the Stanley Neuropathology Consortium - Analysis of
RNAs from the Brains of Individuals with Psychiatric
Diseases (Unpublished) Stanley Neuropathology Laboratory,
Johns Hopkins School of Medicine, Baltimore MD."
BASE COUNT 120 a 73 c 98 g 116 t
ORIGIN

Query Match 7.3%; Score 81.4; DB 38; Length 407;
Best Local Similarity 57.3%; Pred. No. 1.4e-13; Mismatches 207; Conservative 0; Indels 8; Gaps 3;

495 ACTAGCTCAGTACGATACCCGAGACAGTGTATGATGATGATGATGATGATGATGAT 553
382 ATTCCTATCAATTTCACTTCTAGAGACATGATTAATTAATTAATTAATTAATTAATTA 323
554 CGAGTACAGTATCATGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 613
322 ACAGATCAGACTATTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 263
614 ATGTTTAGTGAAT-----TAGCTCACTTGCCCTCAACATCTTGTGATCACTTAT 667
262 ACCTTACCTTTGAAGTGTGACTGCGACCAATTTGGCCCTGCTGCTGATTAATTAAT 203
668 CTTCACTGATTAATGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 727
202 TCCCATTTGATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 144
728 CAGAGTGAAGGATCATATGATGCTTCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 787
143 AGGAAAGACCAATTAAGTCTCTCTCAAAATGACAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 84
788 CTCTGTTAGTGAAGAGTGGCCCTCTGAGACCTCCAGGGTGGTGTGACAGGTACA 847
83 CGCAAGCATTTGTTAGAGTGTATCTTCAACCTCCAGCTGGGATGAGTGGTCTAA 24
848 A 848
23 A 23

RESULT 12
AA542994 523 bp mRNA EST 19-AUG-1997
DEFINITION n15501.s1 NCL CGAP_Ov2 Homo sapiens cDNA clone IMAGE:980713
IMAGE:980713
TITLE Similar to TR:G1020091 G1020091 NEUROPSIN.1, mRNA sequence.
ACCESSION AA542994
NID 92291474
VERSION AA542994.1 GI:2291474

KEYWORDS	EST.
SOURCE	human
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS	1 (bases 1 to 523)
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
COMMENT	Unpublished (1997) On Sep 12, 1996 this sequence version replaced gi:1405040.
<p>Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Christopher A. Moskaluk, M.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: David B. Krizman, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ULNM at: www-bio.11nl.gov/bdrp/Image/Image.html</p>	
<p>Insert Length: 1057 Std Error: 0.00 Seq primer: -21ml3 forward (Amersham) High quality sequence stop: 462.</p>	
<p>location/qualifiers 1..523</p>	
<p>/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:980713" /clone_1lb="NCI_CGAP_Ov2" /sex="female" /tissue_type="Ovary" /lab_host="DH10B" /note="Vector: pAMP10; mRNA made from invasive ovarian tumor, CDNA made by Oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."</p>	
BASE COUNT	112 a 171 c 138 g 102 t
ORIGIN	
<p>Query Match 7.2%; Score 81; DB 35; Length 523; Best Local Similarity 53.7%; Pred. No. 2e-13; Matches 196; Conservative 0; Mismatches 160; Indels 9; Gaps 1;</p>	
0Y	CTCCGACCCCTGGCAGCGGCGACTGTCATGAAACGAATTGTCCTCGCGGCTCT 65
Db	CTTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 111
121	CTCCACCCACCATGGCAGGTCCTCTCATGGCAGCATCAGCTCCACCTCGGAGGCTCT 180
0Y	GGTGATCCGAGTGGGTCGTGACCGCGACACGTCGTTCCAGAACTCTCAGCATGG 125
Db	GGTGTATGAGCGCTGGGTCTACTACGCCCTCAGTGAATGATGATGATGATGATGAT 181
126	GCTGGCCCTGCACAGTCTTGAAGCCGACCAAGCCAGGACGACGATGTGAGGCCAG 185
Db	CCTGGGAGTGATAGCTGGGGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 241
0Y	CCTCTCCGTAGCGGACCCAGAGTACAAAGACCCCTGCTCGCTAAAGACCTCAGTCTAT 245
Db	GAAAGTCATTCGCGCCACCCCGCTACTCCACAGACCCATGTTATGACCTCATCTCTGT 351
246	CAAGTTGACGAATCGTGTCCAGTCTGACACCAATCCGCGAGACACATGCTGTGCA 305
Db	GAAAGTCATTCGCGCCACCCCGCTACTCCACAGACCCATGTTATGACCTCATCTCTGT 352
306	GTCGCCCTACCGGGGAACTCTTGCCCTGTTTCTGGTGGGGTCTGCGGAACGATGC 365
412	CTGGCAACCCCTGGAAACCACTGACGTGTCGCCGGCTGGGGCACTACACAGGCCAGA 471

QY	366	TGTGA	370		
Db	472	TGTGA	476		
RESULT 13					
AA535837					
LOCUS	AA535837				
DEFINITION	AA535837	486 bp	mRNA	EST	21-AUG-1997
	nj79905.s1	NCI-CGAP-Pt10	Homio sapiens	CDNA clone	IMAGE:998744
	similar to	gb:593329	GLANDULAR KALLIKREIN 2	PRECURSOR (HUMAN);	
ACCESSION	AA535837				
NID	AA535837				
VERSION	g2280090				
KEYWORDS	AA535837.1	GI:2280090			
SOURCE	EST.				
ORGANISM	human.				
	Homio sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	Eutheria; Primates; Catarrhini; Homiidae; Homo.				
REFERENCE	1	(bases 1 to 486)			
AUTHORS	NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap.			
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
JOURNAL	Tumor Gene Index				
COMMENT	Unpublished (1997)				
	On Apr 14, 1993 this sequence version replaced gi:315330.				
Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuqui, M.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: David B. Krizman, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www-bio.lnl.gov/bdrp/image/image.html					
Insert Length: 1091 Std Error: 0.00 Seg primer: -40ml3 fwd. ER from Amersham High quality sequence stop: 385. Location/Qualifiers 1. 486 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:998744" /clone_lib="NCI-CGAP_Pt10" /sex="male" /tissue="invasive prostate tumor" /lab_host="DH10B" /note="Organ: prostate; Vector: PAMPI0; mRNA made from invasive prostate tumor. CDNA made by oligo-dr priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Library made by D. Krizman, NIH."					
BASE COUNT 96 a 142 c 145 g 103 t					
ORIGIN					
Query Match 6.88; Score 76.2; DB 35; Length 486; Best Local Similarity 66.88; Pred. No. 5.1e-12; Matches 125; Conservative 0; Mismatches 58; Indels 4; Gaps 1;					
QY	340	GGCTGGGGGTCTGCTGGGCAACGATGCTGTATTGTCATCCAGTCCAGAGCTGGGAAGC	399		
Db	58	GCCCTGTCTGTGGGGTGCACGTGTCCGTGCCCTCATCCAGTCTCGATTGTGGAGGC	117		
QY	400	TGGGAGTGTGAAGAGCTTTCCCAACCTCTGGAGG---GTTTACCATTTTGGCAACTTC	455		
Db	118	TGGGAGTGTGAAGAGCATTTCCCAACCTCTGGAGGCTGTGTACAGTCATGATGGGCA	177		
QY	456	CAGTGCAGAGAGCGTCCCTGCATCCATCCAGTGGGTGCTCACTACTGCATCGATCACC	515		

Db 178 CACTGTGGGGGTCTGCTGCTGACACCCCGAGTGGTCTCACAGCTGCCATTGGCTTAAG 237

QY 516 CGAACA 522
|||
Db 238 AAGATA 244

RESULT 14
AA579026 EST 12-SEP-1997
LOCUS n344026.s1 NCI_CGAP_Prl Homo sapiens cDNA clone IMAGE:915675
DEFINITION similar to gb:539329 GLANDULAR KALLIKREIN 2 PRECURSOR
(HUMAN); contains Ali repetitive element; mRNA sequence.

ACCESSION AA579026 572 bp mRNA
NID 92357210
VERSION AA579026.1 GI:2357210
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS 1 (bases 1 to 572)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407214.

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui,
M.D., Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/bdrip/image/image.html

Insert Length: 921 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 159.
Location/Qualifiers
1. 572
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="915675"
/clone_lib="NCI_CGAP_Prl"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: PAMP10; Site_1: Not1; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected, histologically normal
prostate epithelial cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into PAMP10 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Kitzman."

BASE COUNT 118 a 153 c 181 g 120 t
ORIGIN

Query Match 6.88: Score 76.2; DB 35; Length 572;
Best Local Similarity 66.88; Pred. NO. 5.5e-12;
Matches 123; Conservative 0; Mismatches 58; Indels 4; Gaps 1;

QY 340 GCGTGGGTCTGCTGCGACGATCTGTGATTCATCCAGTCCAGACTGTGGAGGC 399
|||||

Db 16 GCCTTGTCTGTGGGGTGCACACTGTGCCGTCACCCCTCATCCAGTCTCGGATTGTGGAGGC 75

QY 400 TGGGAGTGTGAGAGCTTTTCCCAACCTGGCAGG----GTTGACCATTTCCGCACTTC 455
|||||

Db 76 TGGGAGTGTGAGAGCTTTTCCCAACCTGGCAGGTCGCTGTACACTCATGATGGGCA 135

QY 456 CACTGACAGGAGCTGCTGCTGCTCATCCATCCAGTGGGTGCTCATCTGCTCATCACC 515
|||

Db 136 CACTGTGGGGGTCTGCTGCTGCTGACACCCCGAGTGGTCTCACAGCTGCCATTGGCTTAAG 195

QY 516 CGAACA 522
|||
Db 196 AAGATA 202

RESULT 15
LOCUS T29510 253 bp mRNA EST 06-SEP-1995
DEFINITION EST82448 Human Prostate gland Homo sapiens cDNA 5' end similar to
Kallikrein 1, renal/pancreas/salivary (HT:711), mRNA sequence.
ACCESSION T29510
NID 6611608
VERSION T29510.1 GI:611608
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS 1 (bases 1 to 253)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bull,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chui,M.-W.,
Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S.M.,
Glodde,A., Gnehm,C.L., Hanna,M.C., Heblom,E., Hinkley,P.S.,
Kelley,J.M., Klimk,R.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,
Merrick,J.M., Moreno-Palanges,R.F., McDonald,L.A., Nguyen,D.T.,
Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,
Sauder,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,
Weidman,J.F., Li,T., Bednarek,D.P., Cao,L., Cepeda,M.A.,
Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferlie,A.,
Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,
Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,
Meisner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wing,J., Xu,C.,
Yu,G.-L., Ruben,S.M., Dillon,P.J., Fanoni,M.R., Rosen,C.A.,
Hasselkne,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
Initial Assessment of Human Gene Diversity and Expression Patterns
Based Upon 83 Million Basepairs of cDNA Sequence
Nature 377, 3-174 (1995)
96026280

TITLE
JOURNAL
MEDLINE
COMMENT
CONTACT: Venter, JC
The Institute for Genomic Research
932 Clopper Rd, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: tdbinfo@tdb.tigr.org
For clone availability, additional sequence and expression
information related to this EST, please contact the TIGR Database
(tdbinfo@tdb.tigr.org)
Seq primer: M13 Reverse.
Location/Qualifiers
1. 253
/organism="Homo sapiens"
/db_xref="ATCC (inhost):106507"
/db_xref="taxon:9606"
/clone_lib="Human Prostate gland"
/note="Organ: prostate gland"
BASE COUNT 42 a 68 c 81 g 59 t
ORIGIN

Query Match 6.74: Score 74.8; DB 20; Length 253;

Best Local Similarity 65.8%; Pred. No. 1e-11;
Matches 123; Conservative 0; Mismatches 60; Indels 4; Gaps 1;

```
QY 340 GGCTGGGGTCTGCTGGCGACATGCTGTGATTGCCATCCAGCCAGACTGTGGAGGC 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39 GCTTGTCTGTGGGGTGCACGTGGTGGCCCTCATCCAGTCTGGATTGTGGAGGC 98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 400 TGGGAGTGTGAGAGCTTCCCAACCCCTGGCAGG---GTGTACCATTTGGCACTTC 455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99 TGGGAGTGTGAGAGCATTCACACCCCTGGCAGTGGTGTGTACAGTCATGATGGCA 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 456 CAGTCAGAGAGCTCTCTGCACTCACTGAGTGTCACTACTGCTCATGCAATCACC 515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159 CACTGTGGGGNTTCTCTGTGCACCCCAAGTGGTGTCAACAGCTGCCATTGNTAAG 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 516 CGGACCA 522
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 AAGAAATA 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: September 28, 1999, 11:29:50
Job time: 1871 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 23:36:44 : Search time 2910.9 seconds
(without alignments)
1222.566 Million cell updates/sec

Title: US-09-030-606-177
Sequence: 1 GCGCACTCGAGCCCTGCA.....ATGTTAAAAAAAAAAAAA 1119

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_v1:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_v1:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	360.8	32.2	1347	42	AF113140	Homo sapi

2	270.6	24.2	1140	3	SS076256	U76256 Sus scrofa
3	255.2	22.8	5900	42	AF113141	AF113141 Homo sapi
4	255.2	22.8	4740	42	AF135023	AF135023 Homo sapi
5	255.2	22.8	4385	42	AF148532	AF148532 Homo sapi
6	212.8	19.0	1237	12	AF019979	AF019979 Mus muscu
7	138.4	12.4	199016	9	HS48G12	AL031054 Human DNA
8	134.2	12.0	42179	11	HSU76377	U76377 Human olfac
9	134.2	12.0	39038	11	HSU78308	U78308 Human olfac
10	134.2	12.0	110000	35	AC007194_0	AC007194 Homo sapi
11	134	12.0	61450	9	HS15D7	AL031229 Human DNA
12	129.8	11.6	166832	11	AF070718	AF070718 Homo sapi
13	128.8	11.5	185952	11	AC005906	AC005906 Homo sapi
14	127.8	11.4	208643	34	HS44N10	297197 Homo sapien
15	127	11.3	113693	34	HS41C16	AL049814 Homo sapi
16	125.8	11.2	177471	42	AC007446	AC007446 Homo sapi
17	122.6	11.0	145656	10	HS11K10	AL034449 Homo sapi
18	122.6	11.0	123708	10	HS190A9	AJ006997 Homo sapi
19	121.2	10.8	95930	9	HS332011	298043 Human DNA s
20	120.6	10.8	104961	11	AC004075	AC004075 Homo sapi
21	120	10.7	93727	34	HSB6315	AL049821 Homo sapi
22	119.8	10.7	87552	9	HS217016	AL031771 Human DNA
23	119.8	10.7	132229	9	HS447N6	282211 Human DNA s
24	119.8	10.7	96079	9	HS717L17	AL021706 Human DNA
25	119.8	10.7	93963	11	AF003528	AF003528 Homo sapi
26	119.8	10.7	117896	34	HS287L14	295325 Homo sapien
27	119.6	10.7	180104	11	DJ526N18	AF133462 Homo sapi
28	119	10.6	122157	9	HS4149D17	AL034369 Human DNA
29	118.4	10.6	202763	11	AC005181	AC005181 Homo sapi
30	118.4	10.6	59306	42	AC007423	AC007423 Homo sapi
31	118	10.5	115835	9	HS28C20	AL031119 Human DNA
32	117.2	10.5	165659	11	AF003529	AF003529 Homo sapi
33	117	10.5	46152	11	AC005530	AC005530 Homo sapi
34	117	10.5	47323	11	AC005937	AC005937 Homo sapi
35	116.4	10.4	170282	42	AC006032	AC006032 Homo sapi
36	116.2	10.4	107761	10	HS417022	AL033379 Human DNA
37	115.2	10.3	41025	11	AC003975	AC003975 Human COS
38	115.2	10.3	103916	11	AC004225	AC004225 Homo sapi
39	115	10.3	109343	42	AC007161	AC007161 Homo sapi
40	114.8	10.3	112456	11	AC004614	AC004614 Homo sapi
41	114.8	10.3	180278	35	AC007445	AC007445 Homo sapi
42	113.8	10.2	130381	11	HDAC003003	AC003003 Human Chr
43	113.8	10.2	217873	35	AC006431	AC006431 Homo sapi
44	113.8	10.2	202478	42	AC006560	AC006560 Homo sapi
45	113.6	10.2	129294	9	HS48A11	AL031132 Human DNA

ALIGNMENTS

RESULT 1	AF113140	1347 bp	mRNA	PRI	25-MAR-1999
LOCUS	AF113140				
DEFINITION	Homo sapiens serine protease prostate mRNA, complete cds.				
ACCESSION	AF113140				
NID	54512029				
VERSION	AF113140.1	GT:4512029			
KEYWORDS	human.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS	1 (bases 1 to 1347) Wang, K.				
TITLE	Molecular cloning and characterization of prostate, an androgen-regulated serine protease with prostate-restricted expression				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)				
MEDLINE	99179024				
REFERENCE	2 (bases 1 to 1347)				
AUTHORS	Nelson, P.S., Gan, L., Ferguson, C., Moss, P., Gellinas, R., Hood, L. and Wang, K.				
TITLE	Direct Submission				

JOURNAL Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th St SE,
Bothell, WA 98021, USA
FEATURES
Source location/Qualifiers
1.1347
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13"
1..765
/note="androgen regulated"

CDS

/codon_start=1
/product="serine protease precursor"
/protein_id="AAD21580.1"
/db_xref="PID:94512030"
/db_xref="GI:4512030"
/translation="MTAGNPNWGFILGILGSLAFINGCHINGEDCNPHSOP
QALVMEDELFCGVLVHPQWVLSAHCFQNSYITGLHLSLEADQEPGOMVEASLS
VHPYINPFLANDMLIKLDESVESEDSIRISIASOCPAGNSCLVSGWGLANGR
MPTVQCVNIVSVEVSKLDPYHSMFCAGGHDQKDSNGDSGGPLICNGYLQ
GLVSGKAPCGGVGVYTNLCFTMEIKTVQAS"

polya_site 269 a 489 c 334 g 255 t
BASE COUNT 1327 1332
ORIGIN

Query Match 32.2%; Score 360.8; DB 42; Length 1347;
Best Local Similarity 96.8%; Pred. No. 1.2e-91;

Matches 368; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 2 CGCAGTCGAGCCCTGGCAGGGGCGCACTGTCATGGAAGAAAGATTGCTGCTGGGCG 61
DB 116 CCACACTCGAGCCCTGGCAGGGGCGCACTGTCATGGAAGAAAGATTGCTGCTGGGCG 175
QY 62 TCCTGTGATCGCAGTGGGTGCTGTCAGCCGCACTGTCATGTCCTCAAGTCTTACACCA 121
DB 176 TCCTGTGATCGCAGTGGGTGCTGTCAGCCGCACTGTCATGTCCTCAAGTCTTACACCA 235
QY 122 TCGGGCTGGGCGCTGCAAGTCTTGAAGCCGACCAAGCCAGGAGCCAGATGGTAGG 181
DB 236 TCGGGCTGGGCGCTGCAAGTCTTGAAGCCGACCAAGCCAGGAGCCAGATGGTAGG 295
QY 182 CCAGGCTTCCTGACGCGACCAAGTACACAGACCCTTCTGCTTAACGACCTCATGC 241
DB 296 CCAGGCTTCCTGACGCGACCAAGTACACAGACCCTTCTGCTTAACGACCTCATGC 355
QY 242 TCATCAAGTTGGAGCAATCGGTCCGAGTCTACACATCCGAGCATTCAGATTCCTT 301
DB 356 TCATCAAGTTGGAGCAATCGGTCCGAGTCTACACATCCGAGCATTCAGATTCCTT 415
QY 302 CGCAGTGGCCCTACCGGCGGGAACCTTCTGCTGCTGGGCTGCTGCTGGCAGACG 361
DB 416 CGCAGTGGCCCTACCGGCGGGAACCTTCTGCTGCTGGGCTGCTGCTGGCAGACG 475
QY 362 ATGCTGTGATTGCATTCAG 381
DB 476 GCAGATGCCATCCGTGCTG 495

RESULT 2

SSU76256 1140 bp mRNA MAM 03-JAN-1998
LOCUS Sus scrofa enamel serine proteinase 1 precursor, mRNA,
DEFINITION complete cds.
ACCESSION U76256
NID 92737920
VERSION 076256.1 GI:2737920
KEYWORDS
SOURCE
ORGANISM Sus scrofa
pig.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Suidae; Sus.
REFERENCE 1 (bases 1 to 1140)
AUTHORS Simper,J.P., Fukae,M., Tanabe,T., Yamakoshi,Y., Uchida,T., Xue,J.,

TITLE Margolis,H.C., Shimizu,M., Hu,C.-C. and Bartlett,J.D.
Purification, Characterization and Cloning of Enamel Matrix Serine
Proteinase 1

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1140)
AUTHORS Simper,J.P., Fukae,M., Tanabe,T., Yamakoshi,Y., Uchida,T., Xue,J.,
Margolis,H.C., Shimizu,M., Hu,C.-C. and Bartlett,J.D.

TITLE Direct Submission
JOURNAL Submitted (24-OCT-1996) Pediatric Dentistry, University of Texas
Health Science Center at San Antonio, 7703 Floyd Curl Drive, San
Antonio, TX 78284-7888, USA
location/Qualifiers

FEATURES

source

sig_peptide

CDS

/organism="Sus scrofa"
/db_xref="taxon:9623"
69..140
/note="The signal peptide is cleaved after Ala24 and the
preprotein is secreted into the developing enamel matrix"
69..833
/note="EMSP1"
/codon_start=1
/product="enamel matrix serine proteinase 1 precursor"
/protein_id="AAB94638.1"
/db_xref="PID:92737921"
/db_xref="GI:2737921"
/translation="MTAGNPNWGFILGILGSLAFINGCHINGEDCNPHSOP
QALVMEDELFCGVLVHPQWVLSAHCFQNSYITGLHLSLEADQEPGOMVEASLS
VHPYINPFLANDMLIKLDESVESEDSIRISIASOCPAGNSCLVSGWGLANGR
MPTVQCVNIVSVEVSKLDPYHSMFCAGGHDQKDSNGDSGGPLICNGYLQ
GLVSGKAPCGGVGVYTNLCFTMEIKTVQAS"
159..830
/note="The preprotein is cleaved following His30
activating the serine proteinase; the active protein has
an apparent molecular weight of 34 kDa and a derived
molecular weight of 24 kDa"
/product="enamel matrix serine proteinase 1"
order(177..179,549..551)
/note="disulfide bond"
order(234..236,282..284)
/note="disulfide bond"
join(279..281,414..416,687..689)
/note="encodes catalytic triad"
405..407
/note="encodes potential glycosylation location; yields a
blank cycle during protein sequencing"
order(408..410,789..791)
/note="disulfide bond"
order(510..512,705..707)
/note="disulfide bond"
order(675..677,750..752)
/note="disulfide bond"

BASE COUNT 245 a 366 c 295 g 234 t
ORIGIN

Query Match 24.2%; Score 270.6; DB 3; Length 1140;
Best Local Similarity 84.9%; Pred. No. 3.6e-66;

Matches 303; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 2 CGCAGTCGAGCCCTGGCAGGGGCGCACTGTCATGGAAGAAAGATTGCTGCTGGGCG 61
DB 184 CCACACTCGAGCCCTGGCAGGGGCGCACTGTCATGGAAGAAAGATTGCTGCTGGGCG 243
QY 62 TCCTGTGATCGCAGTGGGTGCTGTCAGCCGCACTGTCATGTCCTCAAGTCTTACACCA 121
DB 244 TCCTGTGATCGCAGTGGGTGCTGTCAGCCGCACTGTCATGTCCTCAAGTCTTACACCA 303
QY 122 TCGGGCTGGGCGCTGCAAGTCTTGAAGCCGACCAAGCCAGGAGCCAGATGGTAGG 181
DB 304 TCGGGCTGGGCGCTGCAAGTCTTGAAGCCGACCAAGCCAGGAGCCAGATGGTAGG 363
QY 182 CCAGGCTTCCTGACGCGACCAAGTACACAGACCCTTCTGCTTAACGACCTCATGC 241
|||||

Query Match	22.88	Score 255.2	DB 42	Length 5900
Best Local Similarity	98.88	Pred. No. 1.1e-61		
Matches 257	Conservative 0	Mismatches 3	Indels 0	Gaps 0
QY 105	CCAAACACTCCACACATCCGCGCTGAGCCCTCACAGTCTTGAGCCCGACCAAGACCCAGG	164		
Db 3313	CCACAGCTCCCTACACATCCGCGCTGAGCCCTCACAGTCTTGAGCCCGACCAAGACCCAGG	3372		
QY 165	GAGCCAGATGGTGGAGGCGACCCCTCTCCGTACGGCCACCCAGATACACAGACCCCTTGGCT	224		
Db 3373	GAGCCAGATGGTGGAGGCGACCCCTCTCCGTACGGCCACCCAGATACACAGACCCCTTGGCT	3432		
QY 225	CGCTACGACCTCATGCTCATCAAGTTGGAGCAATCCGTGCTCCAGTGTGACACATCCG	284		
Db 3433	CGCTACGACCTCATGCTCATCAAGTTGGAGCAATCCGTGCTCCAGTGTGACACATCCG	3492		
QY 285	GAGATCGACGATTCCTTGGAGTCCCTACCGCGGGGGAACCTTGCCCTGTTGGGCTG	344		
Db 3493	GAGATCGACGATTCCTTGGAGTCCCTACCGCGGGGGAACCTTGCCCTGTTGGGCTG	3552		
QY 345	GGGTCTGCTGGCGAAGCATG 364			
Db 3553	GGGTCTGCTGGCGAAGCATG 3572			
RESULT 4				
LOCUS	AF135023	4740 bp	DNA	PRI 20-APR-1999
DEFINITION	Homo sapiens kallikrein-like protein 1 KLR-L1 gene, partial cds.			
ACCESSION	AF135023			
NID	94589272			
VERSION	AF135023.1	GI:4589272		
KEYWORDS	human.			
SOURCE				
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;			
AUTHORS	Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
TITLE	1 (bases 1 to 4740)			
JOURNAL	Yousef,G.M., Luo,L.Y. and Diamandis,E.P.			
REFERENCE	Identification of novel human kallikrein-like genes on chromosome			
AUTHORS	19q13.3-q13.4			
TITLE	Biochem. Biophys. Res. Commun. (1999) In press			
JOURNAL	2 (bases 1 to 4740)			
REFERENCE	Yousef,G.M., Luo,L.Y. and Diamandis,E.P.			
AUTHORS	Direct Submission			
TITLE	Submitted (13-MAR-1999) Pathology and Laboratory Medicine, Mount			
JOURNAL	Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5,			
FEATURES	Canada			
SOURCE	Location/Qualifiers			
	1..4740			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/chromosome="19"			
	/map="19q13.3-q13.4"			
	join(<2263..2425,2847..3097,3181..3317,4588..>4740)			
	join(<2263..2425,2847..3097,3181..3317,4588..4740)			
	/codon_start=3			
	/product="kallikrein-like protein 1 KLR-L1"			
	/protein_id="AAD26424.1"			
	/db_xref="PID:94589273"			
	/translation="SLVSGSCSQTINQEDSPHSOPQWALVNEELFGSGVLYHPQW			
	VLSAARFQNSVTITGLSLHLEADQEPGSDVSRASVSRPEYRNPRLANDLMLIKD			
	YDSSVSPFMSISIASQCPKPNAGSCVSGWGLANGMPVLOCVNVSVSEVGSRL			
	YDLVLPHSFMSIACGGQDQKSDGSGPLICNGYLOGLVSGFKAPCGQVPGVGYTN			
	LCFTEIEKTEVVAS"			
BASE COUNT	1106 a 1275 c 1308 g 1051 t			
ORIGIN				

	Matches	257	Conservative	0	Mismatches	3	Indels	0	Gaps	0
QY	105	CCAGAACTCTCTACCACTTCGGCTGGGCTGGACAGTCTTGAGGCCGACCAAGACCCAG	164							
Db	2841	CCACGCTCTCTACACCATTCGGCTGGGCTGGACAGTCTTGAGGCCGACCAAGAGCCAG	2900							
QY	165	GAGCCAGATGTGTGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACACAGACCCCTTGT	224							
Db	2901	GAGCCAGATGTGTGAGGCCAGCCTCTCCGTACGGCACCCAGAGTACACAGACCCCTTGT	2960							
QY	225	CGCTACGACGCTCATAGCTCATCAATGTGGAGCAATCCGTCCGAGTCCGAAACCATCCG	284							
Db	2961	CGCTACGACGCTCATAGCTCATCAATGTGGAGCAATCCGTCCGAGTCCGAAACCATCCG	3020							
QY	285	GAGCATTCAGCATTTGCTTCGACAGTGCCTACCCGCGGGAATCTTGCTTCGTTGCGTG	344							
Db	3021	GAGCATTCAGCATTTGCTTCGACAGTGCCTACCCGCGGGAATCTTGCTTCGTTGCGTG	3080							
QY	345	GGGTCTGCTGGCGGAACGATG	364							
Db	3081	GGGTCTGCTGGCGGAACGATG	3100							

RESULT	5				
AF148532	AF148532	4385 bp	DNA	PRI	09-JUN-1999
LOCUS					
DEFINITION	Homo sapiens	Kallikrein 4 (KLK4)	gene, complete cds.		
ACCESSION	AF148532				
NID	95020095				
VERSION	AF148532.1	GI:5020095			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 4385)	Stephenson, S.A., Verity, K., Ashworth, L. and Clements, J.A.	Localization of a new prostate specific antigen-related serine

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (03-MAY-1999) School of Life Science, Queensland
University of Technology, George Street, Brisbane, Queensland 4001,
Australia

FEATURES	Location/Qualifiers
source	1. 4385
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="19"
	/map="19q13.3-q13.4"
mrna	join(<551. 611,1875. .2037,2459. .2709,2793. .2929,4200.>4352)
	/gene="KLK4"
	/product="kallikrein 4"
gene	<551. .>4352
	/gene="KLK4"
CDS	join(551. 611,1875. .2037,2459. .2709,2793. .2929,4200. .4352)

BASE COUNT	a	c	g	t
1013	1184	1192	996	

ORIGIN

Query Match	22.8%;	Score 255.2;	DB 42;	Length 4385;
Best Local Similarity	98.8%;	Pred. No. 1e-61;		
Matches 257; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

QY	105	CCACAAATCTCTACACCAATCGGGCTGGGCGCTGCACATCTTGAAGCGCAACAGAGCCAGG	164
Db	2453	CCAAGTCTCTACACCAATCGGGGCTGGGCGCTGCACATCTTGAAGCGCAACAGAGCCAGG	2512
QY	165	GAGCCAATGGTGGAGGCCAGCCTTTCGCTAGGGCACCCAGAGTACAAACAGACCTTCT	224
Db	2513	GAGCCAATGGTGGAGGCCAGCCTTTCGCTAGGGCACCCAGAGTACAAACAGACCTTCT	2572
QY	225	CGCTAACGACCTATGCTCATCAAGTTGGACGAATCCGTCCGAGTCTGACACCATCG	284
Db	2573	CGCTAACGACCTATGCTCATCAAGTTGGACGAATCCGTCCGAGTCTGACACCATCG	2633
QY	285	GAGCATCAGATTGCTTCGCGAGTGCCCTACCGGGGGAACCTTGCTGTTTCGGCTG	344
Db	2633	GAGCATCAGATTGCTTCGCGAGTGCCCTACCGGGGGAACCTTGCTGCTGTTTCGGCTG	2692
QY	345	GGGTCTGCTGGCGAAGCAT	364
Db	2693	GGGTCTGCTGGCGAAGCAT	2712

RESULT	6						
AF019979							
LOCUS	AF019979	1237 bp	-mRNA	ROD	01-JAN-1999		
DEFINITION	Mus musculus enamel matrix serine proteinase 1 precursor, mRNA.						
ACCESSION	AF019979						
MID	94090846						
VERSION	AF019979.1	GI:4090846					
KEYWORDS							
SOURCE	house mouse,						
ORGANISM	Mus musculus						

```

REFERENCE
AUTHORS      1 (bases 1 to 1237)
              Simmer J.
TITLE        Epameli Matrix Serine Proteinase 1 (EMSP1)
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 1237)
AUTHORS      Simmer J.
TITLE        Direct Submission
JOURNAL      Submitted (18-AUG-1997) Pediatric Dentistry, University of Texas
              Health Science Center at San Antonio, 7703 Floyd Curl Drive, San
              Antonio, TX 78284-7888, USA
FEATURES
SOURCE
              1. 1237
                /organism="Mus musculus"
                /strain="Swiss Webster"
                /db_xref="taxon:10090"
                /cell_type="ameloblast"
sig_peptide
CDS

```

misc_feature

repeat_region	/note="LI repeat: matches 5015. .4754 of consensus"
repeat_region	/note="LIM4A repeat: matches 1047. .801 of consensus"
repeat_region	/note="LI repeat: matches 2529. .2371 of consensus"
repeat_region	/note="LI repeat: matches 2203. .1103 of consensus"
repeat_region	/note="MER25 repeat: matches 2136. .1177 of consensus"
repeat_region	/note="MER25 repeat: matches 1031. .582 of consensus"
repeat_region	/note="MER25 repeat: matches 205. .13 of consensus"
repeat_region	/note="MSTA repeat: matches 384. .426 of consensus"
repeat_region	/note="NST-INTERNAL repeat: matches 1. .394 of consensus"
repeat_region	/note="LIMB1 repeat: matches 914. .757 of consensus"
misc_feature	/note="match: STS L41276"
repeat_region	/note="Alusx repeat: matches 1. .299 of consensus"
repeat_region	/note="Alubd repeat: matches 293. .7 of consensus"
repeat_region	/note="THEIB repeat: matches 364. .1 of consensus"
repeat_region	/note="THEIB-INTERNAL repeat: matches 1580. .471 of consensus"
repeat_region	/note="THEIC repeat: matches 44. .3 of consensus"
repeat_region	/note="Alusg repeat: matches 299. .1 of consensus"
repeat_region	/note="MER5b repeat: matches 174. .8 of consensus"
repeat_region	/note="MIR repeat: matches 82. .258 of consensus"
repeat_region	/note="LI repeat: matches 4045. .5390 of consensus"
repeat_region	/note="LI repeat: matches 1. .904 of consensus"
repeat_region	/note="12 copies 2 mer tt 100% conserved"
repeat_region	/note="Aluya5 repeat: matches 301. .1 of consensus"
repeat_region	/note="LIMD1 repeat: matches 971. .806 of consensus"
repeat_region	/note="MR42c repeat: matches 1538. .1485 of consensus"
repeat_region	/note="MR42c repeat: matches 1301. .173 of consensus"
repeat_region	/note="LMC3 repeat: matches 1362. .1054 of consensus"
repeat_region	/note="LIM8 repeat: matches 617. .3 of consensus"
repeat_region	/note="LI repeat: matches 5390. .5158 of consensus"
repeat_region	/note="LIMAs repeat: matches 620. .1016 of consensus"
repeat_region	/note="Aluy repeat: matches 292. .1 of consensus"
repeat_region	/note="LI repeat: matches 4630. .3154 of consensus"
repeat_region	/note="LI repeat: matches 3086. .225 of consensus"
repeat_region	/note="LIMAs repeat: matches 148. .61 of consensus"
repeat_region	/note="LI repeat: matches 5287. .4843 of consensus"
repeat_region	/note="MIR repeat: matches 1. .204 of consensus"
repeat_region	/note="MIR repeat: matches 1. .44801
repeat_region	/note="Alusx repeat: matches 285. .1 of consensus"
repeat_region	/note="MIR repeat: matches 81. .189 of consensus"
repeat_region	/note="MLRIF repeat: matches 541. .379 of consensus"
repeat_region	/note="MIR repeat: matches 262. .120 of consensus"
repeat_region	/note="MIR repeat: matches 64. .137 of consensus"
repeat_region	/note="LIP4 repeat: matches 643. .891 of consensus"
repeat_region	/note="LIP43 repeat: matches 893. .406 of consensus"
repeat_region	/note="MIR repeat: matches 146. .246 of consensus"
misc_feature	/note="match: STS G03807"
repeat_region	/note="MLRID repeat: matches 505. .101 of consensus"
Query Match	
Best Local Similarity 62.2%; Score 138.4; DB 9; Length 199016;	
Matches 270; Conservative 0; Mismatches 156; Indels 8; Gaps 3.	
QY 505	ACTGATCAACCCGGAGACCTGTGATCACTAGACCAGCACCATATTTCGAGTACAGC 564
DB 136146	ATTTCACCTCTAGAACACTGTAAATTAATTAACAAGCCAGAGCTCTACCTGACAGAC 136087
QY 565	TATCATGATTACTGTGTCTACGTGTGTCTATTGTACACCAACGCGCATTTTAGTG 624
DB 136086	TATTCGATTCCTGCTTCCCTGCTGTCTATTATGTAAGTACGCCACCTCTCTTT 136027
QY 625	AATAGCGATCTGAGCCGCACACATCTGTATCCAGTATCCACAGTAT 678
DB 136026	GACAGTGAAGTCTGTGAACCTTGCCACCTGACCTTTGGATCATCATTCATTGTAT 135967
QY 679	TGAGATTTCCCTGCTGATGATGACCATTCGCATTAATTTCTGACCTAGAGGTGAG 738
DB 135966	TAAATTTGAAGTACAGTACGACGACGATTCATGTATGATGACATACAGAAGAC 135907
QY 739	GATCATATAGCTCTTCAAGATGTGTGTCTCCCTCACAAATTCATTCTCTGTGTA 798
DB 135906	AATTAAGGCGCTTCAAGATGACAGTCTCCCTCAAAAATTCATTACAGAAGTAT 135847
QY 799	GTCGAAGTGCAGCCCTGAGCTCCAGGGGTGTCAGATC-ACATGATGATGAT 857
DB 135846	GTCGAAGGATATATCTCTGACCTCTGACGTGGATGATATAGCTTAAGTACTAATG 135787
QY 858	TATGATCGTGTCCCATACCCAAAGCCCTTAATCC-TCATGCTAGTACAGCAGGCG 916
DB 135786	TATTCACCATCCCAATCTCCCAAGCCTTTGATCCCTCTTAACATAAACAAGG 135727
QY 917	AGGTCTAGCATTTTC 930
DB 135726	AGATCAGGCAATTC 135713
RESULT 8	
HSTU76377	HSUT6377 42179 bp DNA PRI 02-JAN-1999
LOCUS Human Olfactory receptor 17-93 (OR17-93) and olfactory receptor	
DEFINITION 17-201 (OR17-201) genes, complete cds.	
ACCESSION U76377	
VERSION g4098234	
KEYWORDS U76377.1 GI:4098234	
SOURCE human.	
ORGANISM Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
Primates; Catarrhini; Hominoidea; Homo.	
REFERENCE 1 (bases 1 to 42179)	
Ferraz,C. and Demallie,J.G.	
Sequence of cosmid ICRF105CH07155 of the olfactory receptor gene	
TITLE	

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 61450)
TITLE Phillips,S.
JOURNAL Direct Submission
Submitted (07-DEC-1998) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquey@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Dec 2, 1998 this sequence version replaced gi:3676174.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
IMPORTANT: This sequence is not the entire insert of clone 15D7. It
may be shorter because we only sequence overlapping sections once,
or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone 3418 (AL021918) is at 61351 in this
sequence. The true right end of clone 97D16 (AL009179) is at 100 in
this sequence. This sequence has been finished according to
sequence map criteria as follows. An attempt is made to resolve all
sequencing problems, such as compressions and repeats, but not
necessarily within known annotated human repeat sequence elements
(e.g. Alu). Where the sequence is ambiguous, there is an annotation
using the 'unsure' feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed in collaboration with the Sanger
Centre chromosome 6 mapping group and Armin Volz & Andreas Ziegler,
David Ruddy, Jeffrey Gruen. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6/>
15D7 is from the library RPC11 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see <http://daccpac.med.buffalo.edu/VECTOR:pcyrc2>.
Location/Qualifiers
1..61450
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="15D7"
/map="p22.1-22.3"
/clone_1lb="RPC11"
381..821
misc_feature
/note="match: GSS AQ213610"
643..950
repeat_region
/note="AluSp repeat: matches 1..309 of consensus"
1312..1385
repeat_region
/note="tRNA-Arg-CGC repeat: matches 1..74 of consensus"
1369..1486
misc_feature
/note="match: STS Z78001"
complement(1494..1797)
/note="match: STS G02930"
1960..2001
repeat_region
/note="21 copies 2 mer ag 76% conserved"
1968..2003
repeat_region
/note="9 copies 4 mer agac 92% conserved"
2803..3107
repeat_region
/note="AluSq repeat: matches 1..308 of consensus"
3292..3366
repeat_region
/note="tRNA-Ile-ATT repeat: matches 1..75 of consensus"
3673..3730
repeat_region
/note="MER66A repeat: matches 284..340 of consensus"
5345..5643
repeat_region
/note="AluX repeat: matches 10..310 of consensus"
5751..5899
repeat_region
/note="FLAM_A repeat: matches 1..125 of consensus"
5903..5954
repeat_region
/note="26 copies 2 mer tg 79% conserved"
5927..5994
repeat_region
/note="7 copies 4 mer tgtg 100% conserved"

misc_feature
6100..6268
/note="match: GSS B55592"
repeat_region
6208..6511
/note="AluSp repeat: matches 1..296 of consensus"
6564..6804
repeat_region
/note="AluY repeat: matches 1..301 of consensus"
7108..7183
repeat_region
/note="tRNA-Phe-TTC repeat: matches 1..73 of consensus"
7205..7250
repeat_region
/note="23 copies 2 mer tt 96% conserved"
7466..7791
repeat_region
/note="AluSp repeat: matches 1..313 of consensus"
8151..8461
repeat_region
/note="AluY repeat: matches 3..304 of consensus"
8462..8579
repeat_region
/note="MER92C repeat: matches 200..311 of consensus"
8580..8664
repeat_region
/note="MER68A repeat: matches 1..84 of consensus"
8785..8920
repeat_region
/note="MER92C repeat: matches 414..552 of consensus"
9471..9769
repeat_region
/note="AluX repeat: matches 1..300 of consensus"
9773..9804
repeat_region
/note="16 copies 2 mer tt 91% conserved"
9882..10170
repeat_region
/note="AluJo repeat: matches 1..297 of consensus"
10186..10368
repeat_region
/note="L1MC2 repeat: matches 6158..6325 of consensus"
10370..10661
repeat_region
/note="AluSq repeat: matches 1..307 of consensus"
10676..10978
repeat_region
/note="AluSq repeat: matches 1..308 of consensus"
10991..11687
repeat_region
/note="L1MC2 repeat: matches 5173..5870 of consensus"
11682..12235
repeat_region
/note="L1MC4 repeat: matches 663..1295 of consensus"
12236..12548
repeat_region
/note="AluX repeat: matches 1..311 of consensus"
12549..12862
repeat_region
/note="L1MC4 repeat: matches 371..663 of consensus"
12937..13246
repeat_region
/note="AluY repeat: matches 1..309 of consensus"
13294..13389
repeat_region
/note="L1MD2 repeat: matches 6242..6330 of consensus"
13390..13687
repeat_region
/note="AluX repeat: matches 12..310 of consensus"
13688..13916
repeat_region
/note="L1MD2 repeat: matches 6031..6242 of consensus"
13917..14217
repeat_region
/note="AluX repeat: matches 1..303 of consensus"
14218..14516
repeat_region
/note="L1MD2 repeat: matches 5660..6031 of consensus"
14522..14691
repeat_region
/note="L1M4 repeat: matches 2357..2493 of consensus"
14692..15025
repeat_region
/note="AluX repeat: matches 1..310 of consensus"
15026..15055
repeat_region
/note="L1M4 repeat: matches 2324..2357 of consensus"
15158..15588
repeat_region
/note="L1M4 repeat: matches 5725..6183 of consensus"
15605..15784
repeat_region
/note="AluB repeat: matches 1..271 of consensus"
15785..15850
repeat_region
/note="L1M4 repeat: matches 5667..5733 of consensus"
15851..16168
repeat_region
/note="AluSc repeat: matches 1..306 of consensus"
16169..16739
repeat_region
/note="L1M4 repeat: matches 5176..5667 of consensus"
16740..17052
repeat_region
/note="AluX repeat: matches 1..312 of consensus"
17053..17519
repeat_region
/note="L1M4 repeat: matches 4696..5176 of consensus"
17541..19073
repeat_region

[illegible][illegible]

QY 631 GCGTACTGGCTCAACCATCTTGATCCAGTATCTCAGTGAATGAGATTCTCG 690
 DB 47341 CCACACTCTGGCCCTGCACTTTGGGATCCCAATATCCACAGCATTTAAATTTGTA 47400
 QY 691 CTTCAGTGTGACCATTCACAC-ATAATTTGTACCTACAGAGTGAGGATCATATAGC 749
 DB 47401 GTTAGTGACAGTGTGCTCTCTGTACATCTGGCATACAGAGCAATTCACAGAGC 47460
 QY 750 TCTTCAAGATGTGTACTCTCCCTCCACAAATTCATCTCTC-CTGTTGTACTGAAGGTG 808
 DB 47461 TCTTCAAGAGGACGAGAGTCTCCCTCCACAAATTCATCTCTC-CTGTTGTACTGAAGGTG 47520
 QY 809 CGCCCTGTGAGCCTCCAGAGGTGTGTGAGGTC-ACAATGATGATGTATGATCTGTG 867
 DB 47521 TGTCTTGTGAGCCTCCAGAGGTGTGTGAGGTC-ACAATGATGATGTATGATCTGTG 47580
 QY 868 TTCCCATTTACCAAGCCTTT-AAATCCCTATGCTCATGATACAGGAGGAGCTGTACCA 926
 DB 47581 TCCCAATCTCCCTTAAGCCTTTGATTTCTCTCTACATTAACCAAGGAGATCAGGCA 47640
 QY 927 TTT 929
 DB 47641 TTT 47643

RESULT 13
 AC005906/c
 LOCUS
 DEFINITION Homo sapiens 1p13.3 BAC RPl11-429A20 (Roswell Park Cancer
 Institute Human BAC Library) complete sequence.
 AC005906
 NID 94165009
 VERSION AC005906.1 GI:4165009
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 185952)
 REFERENCE
 AUTHORS Muzny,D., Arenson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y.,
 Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H.,
 Kondejowski,N., Lau,S., Leal,B., Lee,E., Licharge,O., Liu,W.,
 Logan,O., Lu,J., Marondel,I., Martinez,C., Merscher,S., Miller,A.,
 Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,
 Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shum,C., Simon,M.,
 Vo,Q., Williamson,A., Worley,K.C., Xiang,A.M., Yang,R., Yu,W.,
 Zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A.
 Direct Submission
 TITLE Unpublished
 JOURNAL 2 (bases 1 to 185952)
 REFERENCE
 AUTHORS Worley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (03-NOV-1998) Molecular and Human Genetics, Baylor
 College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 185952)
 REFERENCE
 AUTHORS Worley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (30-DEC-1998) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 4 (bases 1 to 185952)
 REFERENCE
 AUTHORS Worley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (20-JAN-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 5 (bases 1 to 185952)
 REFERENCE
 AUTHORS Worley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (30-JAN-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jan 20, 1999 this sequence version replaced g1:4079596.
 INFORMATION: <http://gc.bcm.tmc.edu:8088/home.html> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
 entire insert of this clone. Overlapping regions of clones are only
 sequenced and submitted once, so the sequence for the remainder of
 the insert may be found in the record for the adjacent clones.
 Overlapping clones are noted at the beginning and end of the
 Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches
 of a local database that includes entries from dbSTS, GDB, and
 local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
 unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
 EST and cDNA sequences. Genes demonstrate at least two exons
 flanked by consensus splice sites that maintained sequence
 continuity across the splice junctions. Sequences that are not
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
 standard of double strand coverage with a minimum of 2 clones and 2
 reads with no ambiguities or 2 chemistries with a minimum of 2
 clones and 3 reads with no ambiguities. If the sequence quality for
 a region does not meet this standard, it will be indicated in the
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
 standards - estimated error rate less than 1 per 10,000 bases.
 Reports of lowest quality individual bases and measures of base
 quality are listed below. Description of the metrics can be found
 at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAR-REPORT-----

Summary Statistics -----

Contig length: 185961
 Phrap values in estimate: 98505
 Average error rate (BCM-Phrap estimate): 5.5169e-07
 Fraction of Phrap values less than 40 : 0.00108624
 Number of N's in consensus : 4
 0

Consensus Changing edits -----

Position Original+Context Edited+Context
 9492 aaaggtatc(t)ttttttttt aaaggtatc(c)ttttttttt
 36770 tggggtggg(n)agcgggagg tggggtggg(g)agcgggagg
 107439 ctggcgctg(n)cgctacact ctggcgctg(a)cgctacact
 109251 gctccgggc(n)ccctacact gctccgggc(c)ccctacact

Bases with BCM-Phrap value < 20-----

Distribution of Quality < 40 Bases -----

Quality	Position	Surrounding Sequence
1001		
901		
801		
701		
601		
501		
401		
301		
201		

```

101
01
-----
5 10 15 20 25 30 35 40
Phrap Value Range

```

Version: 1.01 xfoot.

```

FEATURES
Source
Location/Qualifiers
1.185952
/organism="Homo sapiens"
/db.xref="taxon:9606"
/clone="RPC111-429A20"
/chromosome="12p13.3"
2451..2743
/rpt_family="L1P"
Repeat_region
2734..2959
/rpt_family="L1PA11"
Repeat_region
3054..3118
/rpt_family="L2"
Repeat_region
3332..3426
/rpt_family="MERS3"
Repeat_region
5492..6248
/rpt_family="LIME2"
Repeat_region
6249..6548
/rpt_family="ALUy"
Repeat_region
6549..6742
/rpt_family="LIME2"
Repeat_region
6743..6825
/rpt_family="(CA)n"
Repeat_region
6847..7184
/rpt_family="LIME2"
Repeat_region
7202..7315
/rpt_family="ALUo"
Repeat_region
7316..7421
/rpt_family="L1MA10"
Repeat_region
7951..8224
/rpt_family="L1MC/D"
Repeat_region
9327..9418
/rpt_family="L1M1"
Repeat_region
9499..9765
/rpt_family="ALUy"
Repeat_region
9832..9928
/rpt_family="MSTA"
Repeat_region
10248..10368
/rpt_family="MSTA"
Repeat_region
10374..10668
/rpt_family="ALUSc"
Repeat_region
10669..10901
/rpt_family="MSTA"
Repeat_region
12777..12960
/rpt_family="MTR"
Repeat_region
13055..13459
/rpt_family="LTR20"
Repeat_region
13754..13808
/rpt_family="MTR"
Repeat_region
15214..15393
/rpt_family="ALUJb"
Repeat_region
15863..16122
/rpt_family="L1PB3"
Repeat_region
16203..16361
/rpt_family="MTR"
Repeat_region
16372..16554
/rpt_family="MTR"
Repeat_region
16838..16957
/rpt_family="(CA)n"
Repeat_region
22903..23067
/rpt_family="MERSB"
Repeat_region
23484..23972
/rpt_family="MLT2FA"
Repeat_region
24144..24174
/rpt_family="(CAAAA)n"
Repeat_region

```

```

Repeat_region 24212..25128
/rpt_family="MERVL"
Repeat_region Complement(25147..25701)
/rpt_family="MLT2D"
Repeat_region 25705..29985
/rpt_family="HERVL"
Repeat_region 30191..30523
/rpt_family="MLT2E"

```

Query Match 11.5%; Score 128.8; DB 11; Length 185952;
 Best Local Similarity 68.0%; Pred. No. 1e-25;
 Matches 227; Conservative 0; Mismatches 97; Indels 10; Gaps 3;

```

518 GAACAGTGTGATCACTACCAAGACCAATGCTCCGACGACGATTCATGATGACT 577
|||||
26093 GAAACATGTGATCACTACCAATGCTCCGACGACGATTCATGATGACT 26034
|||||
578 GTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 634
|||||
26033 GCTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 25974
|||||
635 CACTTGCCCTCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 694
|||||
25973 CACTTGCCCTCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 25914
|||||
695 AGTGTGACCATTCCTCA-CATATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 753
|||||
25913 AGTGTGACCATTCCTCA-CATATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 25854
|||||
754 CAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 813
|||||
25853 CATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 25800
|||||
814 TCTGAGCCTCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 847
|||||
25799 TTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 25766
|||||

```

RESULT 14

HS44N10/c

LOCUS Homo sapiens chromosome 12 clone 44N10, WORKING DRAFT SEQUENCE, In
 DEFINITION
 ACCESSION 297197
 NID 93873505
 VERSION 297197.1 GI:3873505
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 208643)
 AUTHORS White,S.
 TITLE Direct Submission
 JOURNAL Submitted (12-NOV-1998) Wellcome Trust Genome Campus, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerresanger.ac.uk Clone requests: clonequest@esanger.ac.uk
 On Nov 16, 1998 this sequence version replaced gi:2225928.

COMMENT

IMPORTANT: This sequence is unfinished and does not necessarily
 represent the correct sequence. Work on the sequence is in progress
 and the release of this data is based on the understanding that the
 sequence may change as work continues. The sequence may be
 contaminated with foreign sequence from E.coli, yeast, vector,
 phage etc. Order of segments is not known; 800 n.s separate
 segments. Unfinished: 44N10 Contig_ID: 03188 Length: 28230 bp
 Unfinished: 44N10 Contig_ID: 03967 Length: 7858 bp Unfinished:
 44N10 Contig_ID: 04590 Length: 16612 bp Unfinished: 44N10
 Contig_ID: 03869 Length: 1065 bp Unfinished: 44N10 Contig_ID: 00878
 Length: 24110 bp Unfinished: 44N10 Contig_ID: 04458 Length: 26725
 bp Unfinished: 44N10 Contig_ID: 04588 Length: 6852 bp Unfinished:
 44N10 Contig_ID: 04631 Length: 9284 bp Unfinished: 44N10 Contig_ID:
 02809 Length: 1644 bp Unfinished: 44N10 Contig_ID: 04439 Length:
 1307 bp Unfinished: 44N10 Contig_ID: 04403 Length: 1278 bp

Cambridgeshire, CB10 1SA, UK. E-mail enquires:

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 09:57:15 : Search time 273 Seconds

(without alignments)
1025.513 Million cell updates/sec

Title: US-09-030-606-177

Perfect score: 1119

Sequence: 1 GCGCACTCGACGACCTGCGCA.....ATGTTAAAAA 1119

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1119	100.0	1119	1	Prostate tumour sp
2	1119	100.0	1119	1	CDNA sequence of p
3	361.4	32.3	1265	1	Prostate tumour sp
4	361.4	32.3	1265	1	CDNA sequence of p
5	352.2	31.5	1167	1	Prostate tumour sp
6	352.2	31.5	1167	1	CDNA sequence of p
7	350.4	31.3	871	1	Human prostate-spe
8	339.8	30.4	1386	1	Human sapiens Tubb1
9	338.8	30.3	1248	1	Prostate tumour sp
10	338.8	30.3	1248	1	CDNA sequence of p
11	338	30.2	402	1	Human secreted pro
12	257.4	23.0	1459	1	Prostate tumour sp
13	257.4	23.0	1459	1	CDNA sequence of p
14	153.8	13.7	234	1	Prostate tumour sp
15	153.8	13.7	234	1	CDNA sequence of p
16	100.4	9.0	1476	1	Human Kallikrein-1
17	93	8.3	907	1	Human Kallikrein-1
18	91	8.1	558	1	BamHI-SacI fragmen
19	91	8.1	379	1	PAK1191 encoding 11
20	91	8.1	555	1	PAK1191 encoding 11
21	91	8.1	309	1	PAK1181 encoding 11
22	91	8.1	309	1	PAK1181 encoding 11
23	89.8	8.0	5406	1	Fragment of clone
24	85	7.6	738	1	Sequence encoding
25	81	7.2	986	1	Human stratum corn
26	81	7.2	1089	1	Human amyloid prec
27	79	7.1	450	1	PAK1181 encoding 11
28	76.4	6.8	832	1	Prostate-specific
29	76.4	6.8	832	1	Prostate-specific
30	76.4	6.8	832	1	Prostate-specific
31	76.4	6.8	832	1	Prostate-specific
32	76.4	6.8	832	1	Prostate-specific
33	76.4	6.8	832	1	Prostate-specific
34	76.4	6.8	832	1	Prostate-specific
35	75.2	6.7	925	1	Prostate-specific
36	73.4	6.6	6139	1	Monkey recombinant
37	73.2	6.5	1466	1	Coding strand of n
38	72.8	6.5	992	1	Prostate specific
39	72.6	6.5	1462	1	Human prostate spe
40	72	6.4	200	1	Human prostate spe
41	71.6	6.4	766	1	mRNA target sequen
42	71.6	6.4	766	1	Prostate-specific
43	71.6	6.4	766	1	Prostate-specific

ALIGNMENTS

RESULT	ID	Score	Query Match	Length	ID	Description
1	V58648	100.0%	100.0%	1119	1	Prostate-specific
2	V58648	100.0%	100.0%	1119	1	Prostate-specific
3	V58648	100.0%	100.0%	1119	1	Prostate-specific
4	V58648	100.0%	100.0%	1119	1	Prostate-specific
5	V58648	100.0%	100.0%	1119	1	Prostate-specific
6	V58648	100.0%	100.0%	1119	1	Prostate-specific
7	V58648	100.0%	100.0%	1119	1	Prostate-specific
8	V58648	100.0%	100.0%	1119	1	Prostate-specific
9	V58648	100.0%	100.0%	1119	1	Prostate-specific
10	V58648	100.0%	100.0%	1119	1	Prostate-specific
11	V58648	100.0%	100.0%	1119	1	Prostate-specific
12	V58648	100.0%	100.0%	1119	1	Prostate-specific
13	V58648	100.0%	100.0%	1119	1	Prostate-specific
14	V58648	100.0%	100.0%	1119	1	Prostate-specific
15	V58648	100.0%	100.0%	1119	1	Prostate-specific
16	V58648	100.0%	100.0%	1119	1	Prostate-specific
17	V58648	100.0%	100.0%	1119	1	Prostate-specific
18	V58648	100.0%	100.0%	1119	1	Prostate-specific
19	V58648	100.0%	100.0%	1119	1	Prostate-specific
20	V58648	100.0%	100.0%	1119	1	Prostate-specific
21	V58648	100.0%	100.0%	1119	1	Prostate-specific
22	V58648	100.0%	100.0%	1119	1	Prostate-specific
23	V58648	100.0%	100.0%	1119	1	Prostate-specific
24	V58648	100.0%	100.0%	1119	1	Prostate-specific
25	V58648	100.0%	100.0%	1119	1	Prostate-specific
26	V58648	100.0%	100.0%	1119	1	Prostate-specific
27	V58648	100.0%	100.0%	1119	1	Prostate-specific
28	V58648	100.0%	100.0%	1119	1	Prostate-specific
29	V58648	100.0%	100.0%	1119	1	Prostate-specific
30	V58648	100.0%	100.0%	1119	1	Prostate-specific
31	V58648	100.0%	100.0%	1119	1	Prostate-specific
32	V58648	100.0%	100.0%	1119	1	Prostate-specific
33	V58648	100.0%	100.0%	1119	1	Prostate-specific
34	V58648	100.0%	100.0%	1119	1	Prostate-specific
35	V58648	100.0%	100.0%	1119	1	Prostate-specific
36	V58648	100.0%	100.0%	1119	1	Prostate-specific
37	V58648	100.0%	100.0%	1119	1	Prostate-specific
38	V58648	100.0%	100.0%	1119	1	Prostate-specific
39	V58648	100.0%	100.0%	1119	1	Prostate-specific
40	V58648	100.0%	100.0%	1119	1	Prostate-specific
41	V58648	100.0%	100.0%	1119	1	Prostate-specific
42	V58648	100.0%	100.0%	1119	1	Prostate-specific
43	V58648	100.0%	100.0%	1119	1	Prostate-specific

D	361	GATCGTGTGATTGGCATTACCGATCCGACAGCTGTGGAGGCGTGGGAGTGTGAAGACTTTCC	420
Q	421	CAACCCGTGGCAGGGTTGTACATTTTCGGCACTTCAGTGCAAAGAGCGTCTGCTGATC	480
D	421	CAACCCGTGGCAGGGTTGTACATTTTCGGCACTTCAGTGCAAAGAGCGTCTGCTGATC	480
Q	481	CTCACTGGGAGCTCAGTACAGTCTACAGCAGACCCGGAACCGAGTGTATCACTATGACAG	540
D	481	CTCACTGGGAGCTCAGTACAGTCTACAGCAGACCCGGAACCGAGTGTATCACTATGACAG	540
Q	541	CACATAGTTCTCCGAAGTCAAGATATCAATGATATAGTACTGTGTAGCTGTCTATTTG	600
D	541	CACATAGTTCTCCGAAGTCAAGATATCAATGATATAGTACTGTGTAGCTGTCTATTTG	600
Q	601	ACTAACATGCCGATGTTTAGTGAAATTAAGCTCACTGGCCTCAACCATCTTGTATC	660
D	601	ACTAACATGCCGATGTTTAGTGAAATTAAGCTCACTGGCCTCAACCATCTTGTATC	660
Q	661	CAGTATACCTTACGATTTAGATTTTCCCTGCTCAGTGTAGCAGCATTTCCACATATTC	720
D	661	CAGTATACCTTACGATTTAGATTTTCCCTGCTCAGTGTAGCAGCATTTCCACATATTC	720
Q	721	TGACCTACAGAGGTGAGGAGATATATAGCTCTTCCAAAGATGTGTACTCCCTCACAA	780
D	721	TGACCTACAGAGGTGAGGAGATATATAGCTCTTCCAAAGATGTGTACTCCCTCACAA	780
Q	781	TTCAATTTCTCCTGTGTAGTGAAGGTGCGCCCTCTGGAGCCTCCAGGGTGGGTGCA	840
D	781	TTCAATTTCTCCTGTGTAGTGAAGGTGCGCCCTCTGGAGCCTCCAGGGTGGGTGCA	840
Q	841	GGTACACATATGATATATATATGTGTGTCCCATTAACCAAGCTTTAATCCCTCATG	900
D	841	GGTACACATATGATATATATATGTGTGTCCCATTAACCAAGCTTTAATCCCTCATG	900
Q	901	CTCAGTACACAGGAGGAGGTGTACATTTCTCAATTAAGTATAGCTGTCCATTACGCA	960
D	901	CTCAGTACACAGGAGGAGGTGTACATTTCTCAATTAAGTATAGCTGTCCATTACGCA	960
Q	961	ACCACTTCAGGACTCTGGATTTCTGCTAGTGAAGTCTCTGATGCTGCTGCTCTGGG	1020
D	961	ACCACTTCAGGACTCTGGATTTCTGCTAGTGAAGTCTCTGATGCTGCTGCTGCTGGG	1020
Q	1021	GAGGTGAGGAGAGGAGGCCCATGTTCAATGGGATCTGTGCAAGTTGTAAACATTAAGTGC	1080
D	1021	GAGGTGAGGAGAGGAGGCCCATGTTCAATGGGATCTGTGCAAGTTGTAAACATTAAGTGC	1080
Q	1081	TTAATTAACAGAGCTGTGATGTTAAAAAAAAAAAAAAA 1119	
D	1081	TTAATTAACAGAGCTGTGATGTTAAAAAAAAAAAAAAA 1119	
RESULT 2			
V61253			
LD	V61253	standard; cDNA; 1119 BP.	
AC	V61253;		
DT	06-JAN-1999	(first entry)	
DE	cDNA sequence of prostate tumour clone P703 splice variant DE14.		
KW	Prostate; cancer; tumour; vaccine; immunogen; clone; ss.		
OS	Homo sapiens.		
PN	W09837093-A2.		
PD	27-AUG-1998.		
PF	25-FEB-1998; U03492.		
PR	09-FEB-1998; US-020956.		
PR	25-FEB-1997; US-806099.		
PR	01-AUG-1997; US-904804.		
PA	(COR1-) CORIXA CORP.		
PI	Dillon DC, Xu J;		
DR	WPI; 98-609886/51.		
DR	P-PSDB: W1873.		
PT	Polypeptides comprising immunogenic portions of prostate proteins -		
PT	used in a vaccine for the treatment of prostate cancer		
PS - Claim 5; Page 108-109; 130pp; English.			

Query Match	100.0%	Score 1119	DB 1	Length 1119
Best Local Similarity	100.0%	Pred. No. 0		
Matches 1119	Conservative 0	Mismatches 0	Indels 0	Gaps 0
OY 1 GGGCACTGCGACACCCCTGGCAGGCGGACACTGGCTCATGAGAAAAGCAATTGTTCTGCTGGGC 60				
Db 1 GGGCACTGCGACCCCTGGCAGGCGGACACTGGCTCATGAGAAAAGCAATTGTTCTGCTGGGC 60				
OY 61 GTCCTGTCATCCGACAGTGGGTGCTGTACGCCGACACACTGTTTCCAGAACTCTACACC 120				
Db 61 GTCCTGTCATCCGAGAGTGGGTGCTGTACGCCGACACACTGTTTCCAGAACTCTACACC 120				
OY 121 ATCGGGGCTGGGCTTCGACAGCTTTTGAGGCCGACCAAGAGCCAGAGGCCAGATGGTGGAG 180				
Db 121 ATCGGGGCTGGGCTTCGACAGCTTTTGAGGCCGACCAAGAGCCAGAGGCCAGATGGTGGAG 180				
OY 181 GCCAGCTCTCCGTAGGGGACCCAGAGTATCAACAGACCCCTTCTGCTTACAGACTCATAG 240				
Db 181 GCCAGCTCTCCGTAGGGGACCCAGAGTATCAACAGACCCCTTCTGCTTACAGACTCATAG 240				
OY 241 CTCATCAAGTTGGAGGAATCCGTGTCCGAGTCTGACACACCATCCGGAGATCAAGACTTGT 300				
Db 241 CTCATCAAGTTGGAGGAATCCGTGTCCGAGTCTGACACACCATCCGGAGATCAAGACTTGT 300				
OY 301 TCGCAGTGGCCCTACCGGGGGGAACCTTTCCTCGTTTTCGGCTGGGGTTCGCTGGGGAGC 360				
Db 301 TCGCAGTGGCCCTACCGGGGGGAACCTTTCCTCGTTTTCGGCTGGGGTTCGCTGGGGAGC 360				
OY 361 GATGCTGTATTGGCATCCAGTCCAGAGTGTGGGAGGCTGGGAGTGTAGAAGCTTTC 420				
Db 361 GATGCTGTATTGGCATCCAGTCCAGAGTGTGGGAGGCTGGGAGTGTAGAAGCTTTC 420				
OY 421 CAACCCGTGGCAGGGTTGATCCATTTTCGGCACTTCCATGTGCAAGAGAGTCTGCTGATC 480				
Db 421 CAACCCGTGGCAGGGTTGATCCATTTTCGGCACTTCCATGTGCAAGAGAGTCTGCTGATC 480				
OY 481 CTCACCTGGCAGGGTTGATCCATTTTCGGCACTTCCATGTGCAAGAGAGTCTGCTGATC 480				
Db 481 CTCACCTGGCAGGGTTGATCCATTTTCGGCACTTCCATGTGCAAGAGAGTCTGCTGATC 480				
OY 541 CACCATAGTCTCCGAAAGTCAGACTATCATGATTACTGTGTGACTGTGCTGTATTGT 600				
Db 541 CACCATAGTCTCCGAAAGTCAGACTATCATGATTACTGTGTGACTGTGCTGTATTGT 600				
OY 601 ACTAACCATGCGAGTGTAGTGAAATTAAGGTCACTTGGCTCAACCATCTTGGTATC 660				
Db 601 ACTAACCATGCGAGTGTAGTGAAATTAAGGTCACTTGGCTCAACCATCTTGGTATC 660				
OY 661 CAGTTATCCTCACTGAATGAGATTTCCTGCTTCACTGATGTGACCATATCCACATAATTTC 720				
Db 661 CAGTTATCCTCACTGAATGAGATTTCCTGCTTCACTGATGTGACCATATCCACATAATTTC 720				
OY 721 TGACCTACAGAGGTGAGGATCATATAGTCTTCAAGAGATGCTGTACTCCCTCAAAA 780				
Db 721 TGACCTACAGAGGTGAGGATCATATAGTCTTCAAGAGATGCTGTACTCCCTCAAAA 780				
OY 781 TTCATTCTCTGTGTAGTGAAGGTGCGCCTCTGAGACCTCCAGAGGTGGGTGTGCA 840				
Db 781 TTCATTCTCTGTGTAGTGAAGGTGCGCCTCTGAGACCTCCAGAGGTGGGTGTGCA 840				
OY 841 GGGCAAAATGATGAATGATGATGCGTGTCCATTTCCAAACCAAGCCTTTAATCCCTCATG 900				
Db 841 GGGCAAAATGATGAATGATGATGCGTGTCCATTTCCAAACCAAGCCTTTAATCCCTCATG 900				


```

V58647
ID V58647 standard; cDNA; 1167 BP.
AC V58647;
DE 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE13.
KM Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy; ss.
OS Homo sapiens.
PI Key
PI CDS Location/Qualifiers
FT CDS 28..645
FT tag= a
FT WO9837418-A2.
PD 27-AUG-1998.
PD 25-FEB-1998: U03690.
PR 09-FEB-1998: US-904809.
PR 25-FEB-1997: US-806586.
PR 01-AUG-1997: US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
PI WPI: 98-480805/41.
DR P-PSDB: W69388.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 115; 141p; English.
CC This sequence represents a human prostate tumour specific gene, and
CC can be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;

Query Match
Best Local Similarity 31.5%; Score 352.2; DB 1; Length 1167;
Matches 360; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 8 CCGAGCCCTGGCAGCGGCGACATGTCATGGAAGAAAGATTTGCTGCGGGCGTCTGG 67
DB 2 CCGAGCCCTGGCAGCGGCGACATGTCATGGAAGAAAGATTTGCTGCGGGCGTCTGG 61
QY 68 TGCATCCGAGTGGGTGCTGTGACGCGGACACTGTTCCAGAACTCTACACATGGGGC 127
DB 62 TGCATCCGAGTGGGTGCTGTGACGCGGACACTGTTCCAGAACTCTACACATGGGGC 121
QY 128 TGGGCGCTGCACAGTCTTGAAGCGCCAGCAAGAGCCAGAGGAGCATGTTGAGGCGCAGCC 187
DB 122 TGGGCGCTGCACAGTCTTGAAGCGCCAGCAAGAGCCAGAGGAGCATGTTGAGGCGCAGCC 181
QY 188 TCTCCGTACGAGCACCAGAGTACACAGACCCCTGCTCGCTAAAGACTTATGCTCATCA 247
DB 182 TCTCCGTACGAGCACCAGAGTACACAGACCCCTGCTCGCTAAAGACTTATGCTCATCA 241
QY 248 AGTTGACGAATCCGTGCTGCGAGTGTACACCATCCGGAGCATACACATTCCTCCAGT 307
DB 242 AGTTGACGAATCCGTGCTGCGAGTGTACACCATCCGGAGCATACACATTCCTCCAGT 301
QY 308 GCCCTACCGCGGGGAACTCTTGCTGCTTGGCTGGGGTGTGTCGCGAAGATGCTG 367
DB 302 GCCCTACCGCGGGGAACTCTTGCTGCTTGGCTGGGGTGTGTCGCGAAGATGCTG 361
QY 368 TGATTGCAATCCAG 381
DB 362 TGCCCTACCGTGTG 375

RESULT 6
ID V61252 standard; cDNA; 1167 BP.
AC V61252;

```

```

DE 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE13.
KM Prostate; Cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PI Key
PI CDS Location/Qualifiers
FT CDS 28..645
FT tag= a
FT WO9837093-A2.
PD 27-AUG-1998.
PD 25-FEB-1998: U03492.
PR 09-FEB-1998: US-020956.
PR 25-FEB-1997: US-806099.
PR 01-AUG-1997: US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
PI WPI: 98-609886/51.
DR P-PSDB: W71782.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 107; 130p; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;

Query Match
Best Local Similarity 31.5%; Score 352.2; DB 1; Length 1167;
Matches 360; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 8 CCGAGCCCTGGCAGCGGCGACATGTCATGGAAGAAAGATTTGCTGCGGGCGTCTGG 67
DB 2 CCGAGCCCTGGCAGCGGCGACATGTCATGGAAGAAAGATTTGCTGCGGGCGTCTGG 61
QY 68 TGCATCCGAGTGGGTGCTGTGACGCGGACACTGTTCCAGAACTCTACACATGGGGC 127
DB 62 TGCATCCGAGTGGGTGCTGTGACGCGGACACTGTTCCAGAACTCTACACATGGGGC 121
QY 128 TGGGCGCTGCACAGTCTTGAAGCGCCAGCAAGAGCCAGAGGAGCATGTTGAGGCGCAGCC 187
DB 122 TGGGCGCTGCACAGTCTTGAAGCGCCAGCAAGAGCCAGAGGAGCATGTTGAGGCGCAGCC 181
QY 188 TCTCCGTACGAGCACCAGAGTACACAGACCCCTGCTGCTAAAGACTTATGCTCATCA 247
DB 182 TCTCCGTACGAGCACCAGAGTACACAGACCCCTGCTGCTAAAGACTTATGCTCATCA 241
QY 248 AGTTGACGAATCCGTGCTGCGAGTGTACACCATCCGGAGCATACACATTCCTCCAGT 307
DB 242 AGTTGACGAATCCGTGCTGCGAGTGTACACCATCCGGAGCATACACATTCCTCCAGT 301
QY 308 GCCCTACCGCGGGGAACTCTTGCTGCTTGGCTGGGGTGTGTCGCGAAGATGCTG 367
DB 302 GCCCTACCGCGGGGAACTCTTGCTGCTTGGCTGGGGTGTGTCGCGAAGATGCTG 361
QY 368 TGATTGCAATCCAG 381
DB 362 TGCCCTACCGTGTG 375

RESULT 7
ID V37495 standard; DNA; 871 BP.
AC V37495;
DE 07-SEP-1998 (first entry)
DE Human prostate-specific kallikrein (HPSK) encoding DNA.
DE Prostate-specific kallikrein; HPSK; prostate carcinoma; human;
KW benign prostate hyperplasia; diagnosis; drug screening; PSK; ss.
OS Homo sapiens.
PI Key
PI CDS Location/Qualifiers
FT CDS 31..777
FT tag= a
FT trans_except= (pos:367..369, aa:Xaa)
FT trans_except= (pos:412..414, aa:Xaa)

```


KW Prostate tumour specific gene; human; prostate cancer; detection;
 KM therapy; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 217..696
 PI /tag= a
 PD WO9837418-A2.
 PE 27-AUG-1998.
 PF 25-FEB-1998; U03690.
 PR 09-FEB-1998; US-904809.
 PR 25-FEB-1997; US-806596.
 PR 01-AUG-1997; US-904809.
 PA (CORI-) CORIXA CORP.
 PI Dillon DC, Xu J;
 DR WPI: 98-480805/41.
 DR P-PSDB: W69387.
 PT Novel human prostate specific tumour protein and fragments - useful
 PT for detecting and treating prostate cancers
 CC Claim 1; Page 112; 141pp; English.
 CC This sequence represents a human prostate tumour specific gene, and can
 be used in the method of the invention. The method is for detecting
 CC prostate cancer comprises contacting a biological sample with an agent
 CC able to bind an immunogenic portion of a prostate protein (such as
 CC encoded by this sequence). An antibody which binds to an immunogenic
 CC portion of the prostate protein, and the method can be used to detect,
 CC monitor progression of, or treat prostate cancers. The antibody may
 CC also be conjugated to a therapeutic agent for use in therapy of prostate
 CC cancers.
 SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;
 Query Match 30.3%; Score 338.8; DB 1; Length 1248;
 Best Local Similarity 93.9%; Pred. No. 3.4e-91;
 Matches 368; Conservative 0; Mismatches 12; Indels 12; Gaps 1;
 KY 2 CGCAGCTCGAGCCCTGGCGAGCGGCACTGTATGAGAAAGCAATTGTTCTGCTGGGCG 61
 DB CGCAGCTCGAGCCCTGGCGAGCGGCACTGTATGAGAAAGCAATTGTTCTGCTGGGCG 94
 KY 62 TCCTGTGATCGGAGTGGGCTGTGACGCGACACTGTTTCCAGAA----- 111
 DB TCCTGTGATCGGAGTGGGCTGTGACGCGACACTGTTTCCAGAAAGTGTGAGCA 154
 KY 111 -CTCTACACCATCGGGCTGGGCTGACAGTCTTGAGGCGGAGAGAGAGAGAGAGCC 169
 DB GCTCTACACCATCGGGCTGGGCTGACAGTCTTGAGGCGGAGAGAGAGAGAGAGAGCC 214
 KY 170 AGATGTGAGGCGGAGCTCTCGTACGAGCAGCCAGATGACAAAGACCTTGTGCTA 229
 DB AGATGTGAGGCGGAGCTCTCGTACGAGCAGCCAGATGACAAAGACCTTGTGCTA 274
 KY 215 AGATGTGAGGCGGAGCTCTCGTACGAGCAGCCAGATGACAAAGACCTTGTGCTA 274
 DB AGATGTGAGGCGGAGCTCTCGTACGAGCAGCCAGATGACAAAGACCTTGTGCTA 274
 KY 230 AGACCTCATGCTCATCAAGTTGGAGCAATCGTGTGAGTGTGACACCATCGGAGCA 289
 DB AGACCTCATGCTCATCAAGTTGGAGCAATCGTGTGAGTGTGACACCATCGGAGCA 334
 KY 275 ACACCTCATGCTCATCAAGTTGGAGCAATCGTGTGAGTGTGACACCATCGGAGCA 334
 DB ACACCTCATGCTCATCAAGTTGGAGCAATCGTGTGAGTGTGACACCATCGGAGCA 334
 KY 290 TCAGCAATGCTTGGAGTGGGCTGACAGGAGGAGACTTCTGCTGTTGCTGGGCTC 349
 DB TCAGCAATGCTTGGAGTGGGCTGACAGGAGGAGACTTCTGCTGTTGCTGGGCTC 394
 KY 335 TCAGCAATGCTTGGAGTGGGCTGACAGGAGGAGACTTCTGCTGTTGCTGGGCTC 394
 DB TCAGCAATGCTTGGAGTGGGCTGACAGGAGGAGACTTCTGCTGTTGCTGGGCTC 394
 KY 350 TCCTGGCGAAGCATGCTGTGATTCATCCAG 381
 DB TCCTGGCGAAGCATGCTGTGATTCATCCAG 426
 RESULT 10
 V61249 standard; cDNA; 1248 BP.
 AC V61249;
 DT 06-JAN-1999 (first entry)
 DE cDNA sequence of prostate tumour clone P703 splice variant DEL.
 KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
 OS Homo sapiens.
 PN WO9837093-A2.

PD 27-AUG-1998.
 PE 25-FEB-1998; U03492.
 PR 09-FEB-1998; US-020956.
 PR 25-FEB-1997; US-806089.
 PR 01-AUG-1997; US-904804.
 PA (CORI-) CORIXA CORP.
 PI Dillon DC, Xu J;
 DR WPI: 98-609886/51.
 DR P-PSDB: W71871.
 PT Polypeptides comprising immunogenic portions of prostate proteins -
 PT used in a vaccine for the treatment of prostate cancer
 CC Claim 3; Page 104; 130pp; English.
 CC The present sequence is a new DNA which encodes an immunogenic portion
 CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
 CC can be used as a vaccine for the treatment of prostate cancer. The DNA
 CC was identified by analysis of a subtracted cDNA library obtained by
 CC subtracting a prostate tumour cDNA expression library with a normal
 CC tissue cDNA library.
 SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;
 Query Match 30.3%; Score 338.8; DB 1; Length 1248;
 Best Local Similarity 93.9%; Pred. No. 3.4e-91;
 Matches 368; Conservative 0; Mismatches 12; Indels 12; Gaps 1;
 KY 2 CGCAGCTCGAGCCCTGGCGAGCGGCACTGTATGAGAAAGCAATTGTTCTGCTGGGCG 61
 DB CGCAGCTCGAGCCCTGGCGAGCGGCACTGTATGAGAAAGCAATTGTTCTGCTGGGCG 94
 KY 62 TCCTGTGATCGGAGTGGGCTGTGACGCGACACTGTTTCCAGAA----- 111
 DB TCCTGTGATCGGAGTGGGCTGTGACGCGACACTGTTTCCAGAAAGTGTGAGCA 154
 KY 111 -CTCTACACCATCGGGCTGGGCTGACAGTCTTGAGGCGGAGAGAGAGAGAGAGAGCC 169
 DB GCTCTACACCATCGGGCTGGGCTGACAGTCTTGAGGCGGAGAGAGAGAGAGAGAGCC 214
 KY 155 GCTCTACACCATCGGGCTGGGCTGACAGTCTTGAGGCGGAGAGAGAGAGAGAGAGCC 214
 DB GCTCTACACCATCGGGCTGGGCTGACAGTCTTGAGGCGGAGAGAGAGAGAGAGAGCC 214
 KY 170 AGATGTGAGGCGGAGCTCTCGTACGAGCAGCCAGATGACAAAGACCTTGTGCTA 229
 DB AGATGTGAGGCGGAGCTCTCGTACGAGCAGCCAGATGACAAAGACCTTGTGCTA 274
 KY 215 AGATGTGAGGCGGAGCTCTCGTACGAGCAGCCAGATGACAAAGACCTTGTGCTA 274
 DB AGATGTGAGGCGGAGCTCTCGTACGAGCAGCCAGATGACAAAGACCTTGTGCTA 274
 KY 230 AGACCTCATGCTCATCAAGTTGGAGCAATCGTGTGAGTGTGACACCATCGGAGCA 289
 DB AGACCTCATGCTCATCAAGTTGGAGCAATCGTGTGAGTGTGACACCATCGGAGCA 334
 KY 275 ACACCTCATGCTCATCAAGTTGGAGCAATCGTGTGAGTGTGACACCATCGGAGCA 334
 DB ACACCTCATGCTCATCAAGTTGGAGCAATCGTGTGAGTGTGACACCATCGGAGCA 334
 KY 290 TCAGCAATGCTTGGAGTGGGCTGACAGGAGGAGACTTCTGCTGTTGCTGGGCTC 349
 DB TCAGCAATGCTTGGAGTGGGCTGACAGGAGGAGACTTCTGCTGTTGCTGGGCTC 394
 KY 335 TCAGCAATGCTTGGAGTGGGCTGACAGGAGGAGACTTCTGCTGTTGCTGGGCTC 394
 DB TCAGCAATGCTTGGAGTGGGCTGACAGGAGGAGACTTCTGCTGTTGCTGGGCTC 394
 KY 350 TCCTGGCGAAGCATGCTGTGATTCATCCAG 381
 DB TCCTGGCGAAGCATGCTGTGATTCATCCAG 426
 RESULT 11
 X41114 standard; cDNA; 402 BP.
 AC X41114;
 DT 17-JUN-1999 (first entry)
 DE Human secreted protein 5' EST SEQ ID NO:58.
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW differentiation; haematopoiesis regulation; tissue growth regulation;
 KW reproductive hormone regulation; chemotactic; chemokine; haemostatic;
 KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
 OS Homo sapiens.
 PN WO9906548-A2.
 PD 11-FEB-1999.
 PF 31-JUL-1998; IB1222.
 PR 01-AUG-1997; US-905135.
 PA (GENT) GENSET.
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;

DR WPI: 99-153778/13.
DR P-PSDB: Y12281.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
PS Claim 1: Page 205; 824bp; English.
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, hematopoiesis regulating activity, tissue growth regulation
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 402 BP; 80 A; 125 C; 121 G; 75 T;

Query Match 30.2%; Score 338; DB 1; Length 402;

Best Local Similarity 96.8%; Pred. No. 3.4e-91; Matches 360; Conservative 0; Mismatches 0; Indels 12; Gaps 1;

QY 2 CGCACTCGAGCCCTGGCAGCGCGCACTGTCATGGAACGAATGTCCTCGGCGG 61
DB 23 CGCACTCGAGCCCTGGCAGCGCGCACTGTCATGGAACGAATGTCCTCGGCGG 82
QY 62 TCCTGTCATCGCAGTGGGTGCTGTCAGCCGACACTGTTCCAGAA----- 111
DB 83 TCCTGTCATCGCAGTGGGTGCTGTCAGCCGACACTGTTCCAGAAAGTGAATKACA 142
QY 111 -CTCTACACCATGGGGGCGGCGCTGCAAGTCTGAGGCGGACCAAGAGCCAGGAGCC 169
DB 143 GCTCTTACACCATGGGGGCGGCGCTGCAAGTCTGAGGCGGACCAAGAGCCAGGAGCC 202
QY 170 AGATGTCGAGGCGCAGCCCTCCGTAAGGCGCAGAGTAACAAGACCTTGTCTGCTA 229
DB 203 AGATGTCGAGGCGCAGCCCTCCGTAAGGCGCAGAGTAACAAGACCTTGTCTGCTA 262
QY 230 ACGACCTCATGCTCATCAAGTTGAGCAATCCGTCTCGAGTCTGACACCATCCGAGCA 289
DB 263 ACGACCTCATGCTCATCAAGTTGAGCAATCCGTCTCGAGTCTGACACCATCCGAGCA 322
QY 290 TCAGATTCGCTTCGAGAGTCCCTACCGGGGGAATCTTGGCTTCTGCTGGGTC 349
DB 323 TCAGATTCGCTTCGAGAGTCCCTACCGGGGGAATCTTGGCTTCTGCTGGGTC 382
QY 350 TGCTGGCGAAGC 361
DB 383 TGCTGGCGAAGC 394

RESULT 12

ID V58646 standard; cDNA; 1459 BP.
AC V58646:
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE6.
KW Prostate tumour specific gene; human; prostate cancer; detection;
OS Homo sapiens.
PN M09837/18-A2.
PD 27-AUG-1998.
PR 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.

PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J.
DR WPI: 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1: Page 114; 141bp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprising contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Query Match 23.0%; Score 257.4; DB 1; Length 1459;

Best Local Similarity 97.8%; Pred. No. 6.3e-67; Matches 261; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 98 ACTGTTCCAGAACTCTACACCAATCGGCTGGCTGCGACAGTCTTGAAGCGGACCAAG 157
DB 25 AGTAGTCGACAGCTCTACACCAATCGGCTGGCTGCGACAGTCTTGAAGCGGACCAAG 84
QY 158 AGCCAGGAGCCAGATGTTGAGGCGCAGCTCTCGTAAGCGCACCAGAGTAACAAGAC 217
DB 85 AGCCAGGAGCCAGATGTTGAGGCGCAGCTCTCGTAAGCGCACCAGAGTAACAAGAC 144
QY 218 CTTGCTGCGTAAGACCTCATGCTCATGAGTGAAGAAATCGGTGCGAGTCTGACA 277
DB 145 CTTGCTGCGTAAGACCTCATGCTCATGAGTGAAGAAATCGGTGCGAGTCTGACA 204
QY 278 CCATCCGAGGATGAGATGCTTCTGCGAGTCCCTACCGCGGGAATCTTGCCTCGTTT 337
DB 205 CCATCCGAGGATGAGATGCTTCTGCGAGTCCCTACCGCGGGAATCTTGCCTCGTTT 264
QY 338 CTGGCTGGGGTCTGCTGGCGAAGCATG 364
DB 265 CTGGCTGGGGTCTGCTGGCGAAGCATG 291

RESULT 13

ID V61251 standard; cDNA; 1459 BP.
AC V61251:
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE6.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN M09837/093-A2.
PD 27-AUG-1998.
PR 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J.
DR WPI: 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3: Page 106; 130bp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Wed Sep 29 14:27:38 1999

us-09-030-606-177_1.rng

Page 9

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 07:30:09 ; Search time 165.01 Seconds
(Without alignments)
621.696 Million cell updates/sec

Title: US-09-030-606-177
Perfect score: 1119
Sequence: 1 GCCGACTCGCAGCCCTGCA.....ATGTTAAAAA 1119

Scoring table: IDENTITY_NUC

Searched: 176461 seqs, 45838279 residues

Database: Issued Patents, NA:*

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/5C.COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/5D.COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCUS9.COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	350.4	31.3	871	3	US-08-744-026-2
2	81	7.2	986	4	US-08-557-146-1
3	81	7.2	1089	5	PCT-US96-04294-1
4	81	7.2	1089	5	PCT-US96-04294-3
5	76.4	6.8	832	5	PCT-US95-06157-5
6	72.8	6.5	992	2	US-08-358-782D-13
7	72.6	6.5	1462	2	US-08-358-782D-14
8	72	6.4	200	3	US-08-931-981A-1
9	71.6	6.4	766	5	PCT-US95-06157-9
10	68	6.1	1729	4	US-08-844-024-1
11	63.6	5.7	760	5	PCT-US95-06157-7
12	62.4	5.6	833	4	US-08-790-137-2
13	58.4	5.2	2259	4	US-08-845-998-3
14	56.8	5.1	2218	4	US-08-845-998-5
15	50.8	4.5	957	3	US-08-684-862-11
16	49.6	4.4	1454	3	US-08-467-155A-2
17	49.6	4.4	1454	4	US-08-628-198-2
18	49.6	4.4	1454	5	PCT-US96-07343-2
19	46	4.1	840	3	US-08-684-862-12
20	45.6	4.1	732	3	US-08-361-395-2
21	45	4.0	1112	3	US-08-454-720A-42
22	40.8	3.6	925	1	US-07-990-301A-3
23	40.4	3.6	7218	2	US-08-232-463-14
24	40.2	3.6	734	3	US-08-650-129-1
25	40.2	3.6	821	3	US-08-650-129-2
26	40.2	3.6	866	3	US-08-650-129-3
27	38.4	3.4	1096	3	US-08-684-862-8
28	36.4	3.3	1554	3	US-08-465-486-1
29	36	3.2	970	1	US-08-148-910-3
30	36	3.2	2033	1	US-08-148-910-14
31	36	3.2	970	2	US-08-448-937A-3
32	36	3.2	2033	2	US-08-448-937A-14
33	35.8	3.2	959	3	US-08-568-031-1
34	35.8	3.2	959	4	US-08-568-031-1
35	35.4	3.2	2743	2	US-08-396-479B-3
36	35.4	3.2	2743	2	US-08-818-823-3
37	35.4	3.2	2749	4	US-08-124-981A-1

ALIGNMENTS

38	35.4	3.2	2853	5	PCT-US94-07297-36	Sequence 36, Appl
39	35	3.1	807	2	US-08-270-584A-1	Sequence 1, Appl
40	35	3.1	807	4	US-08-765-192-1	Sequence 1, Appl
41	34.8	3.1	867	1	US-07-990-301A-1	Sequence 1, Appl
42	34.2	3.1	1333	3	US-08-684-862-9	Sequence 9, Appl
43	33.8	3.0	681	1	US-07-929-198-1	Sequence 1, Appl
44	33.8	3.0	681	1	US-07-929-198-3	Sequence 3, Appl
45	33.8	3.0	681	1	US-07-929-198-5	Sequence 5, Appl

RESULT 1
US-08-744-026-2
Sequence 2, Application US/08744026

Patent No. 5786148

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

APPLICANT: Goll, Surya K.

TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESSES:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: US

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/744,026

FILING DATE: Herewith

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0154 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ. ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 871 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY:

CLONE: Consensus

US-08-744-026-2

Query Match 31.3%; Score 350.4; DB 3; Length 871;
Best Local Similarity 95.0%; Pred. No. 5.6e-103;

Matches 360; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY	3	GCATCGCAGCCCTGCGAGCGGACATGTCATGGAAGCAATGTTCTGCGGCGT	62
DB	129	GCATTCGACGCGCCGCGAGCGGACATGTCATGGAAGCAATGTTCTGCGGCGT	188
QY	63	CGTGTGATCGGAGGTGCTGTCAGCGGACACTGTTCCAGACTCTACACCAT	122
DB	189	CGTGTGATCGGAGGTGCTGTCAGCGGACACTGTTCCAGACTCTACACCAT	248

Oy	123	CGGGCTGGGCGCTGCACACTCTTTGAGGCCGACCAAGACAGAGAGCCAGATGGTGGAGGC	182
Db	249	CGGGCTGGGCGCTGCACACTCTTTGAGGCCGACCAAGAGAGCCAGATGGTGGAGGC	308
Oy	183	CAGCTCTCCGTACGGCACCAGAGTACACAGACCCTTGCTGCTACAGACCTCATGCT	242
Db	309	CAGCTCTCCGTACGGCACCAGAGTACACAGACCCTTGCTGCTACAGACCTCATGNT	368
Oy	243	CATCAAGTTGGAGCAATCCGTGTCGAGCTGCACACCATTCCGGAGCATACGACATTGCTTC	302
Db	369	CATCAAGTTGGAGCAATCCGTGTCGAGCTGCACACCATTCCGGAGNATACACATTGTTTC	428
Oy	303	GCAGTGGCCCTACCGCGGGGAACTCTTGCTGCTTTCGCTGCGGTGCTGCGGAACGA	363
Db	429	GCAGTGGCCCTACCGCGGGGAACTTTTGCTGCTTTCGCTGCGGTGCTGCGGAACGG	488
Oy	363	TGCTGTGAATGGCATCCAG	381
Db	489	CAGAAATGCTTACCGTCTG	507

```

RESULT 2
US-08-557-146-1
Sequence 1, Application US/08557146
Patent No. 5834290
GENERAL INFORMATION:
APPLICANT: Egelund, Torbjorn
APPLICANT: Hansson, Lennart
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
TITLE OF INVENTION: Enzyme (SCCE)
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: White & Case, Patent Department
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2787
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (E0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,146
FILING DATE: 14-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Steiner, Richard J.
REGISTRATION NUMBER: 35,372
REFERENCE/DOCKET NUMBER: 1103326-181
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 354-8783
TELEFAX: (212) 354-8113
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 986 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 25..786
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 25..90
FEATURE:
NAME/KEY: mat_peptide

```

LOCATION: 112..783
US-08-557-146-1

Query Match	7.2%;	Score 81;	DB 4;	Length 986;
Best Local Similarity	53.7%;	Pred. No. 1.8e-16;		
Matches 196;	Conservative	0;	Mismatches 160;	Indels 9;
				Gaps 1.

QY	6	CTCGAGCCCTGAGCAGCGGCGACACGGGTCATGTGAAAGCAATGTGCTCTCGGGCGTCT	65
Db	141	CTCCACACCCATTGGCAGGAGGGCCCTGCTACAGTGGCAATCAGCTCCACATCGGGAGGCGTCT	200
QY	66	GGTCGATCCCGACATGGGTGCTGTATAGCCGCACACTGTTTCCAGAACTCTTACACCATGG	125
Db	201	GGTCATATAGCGCTGAGGGGCTCTACATGCCGCCCACTGCAGATGAAATAGTACACCTGCA	260
QY	126	GCTGGGCGCTGCAAGTCTTTAGAGGCCGACCAAGACCCAGGGGAGCCAGATGGTGGAGGGCAG	185
Db	261	CCTGGGCGATGATACGGCTGGGCGACAGAGAG-----CTTAGAGGATCAAGGGCTTC	311
QY	186	CCTTCCTGACGAGACCCAGAGCTCAACAGAGACCCCTTGCTGCTATACACCTCATGCTCAT	245
Db	312	GAGTCATTTCCGGCCACCCCGGCTACTCCACACAAACCATGTTAATGACCTCATGCTGCT	377
QY	246	CAAGTTGGACGAATTCGCTGTCCGAGTCTGACACCATCCGGAGCATTCGATGCTTGCGCA	305
Db	372	GAAGTCTCAATGAGCCAGGCGCAGAGGCTGTATCCATGSTGAAGAAATCAGAGGCTGCCCTCCG	433
QY	306	GTCGCCCTTACCGGGGGAATCTTGCTGCTGTTTCTGCGCTGGGGTCTGCTGGCGAAGCATGC	365
Db	432	CTGCGAACCCTTGAGAACCATCTTACTGTCTCCGGCTGGGGGCACACTACACGAGACCCAGA	491
QY	366	TGTGA 370	
Db	492	TGTGA 496	

RESULT 3
PCT-US96-04294-1/c
Sequence 1, Application PC/TUS9604294
GENERAL INFORMATION:
APPLICANT: Dixon, Eric P.
APPLICANT: Johnstone, Edward M.
APPLICANT: Iattle, Shella P.
TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
TITLE OF INVENTION: RELATED NUCLEIC ACIDS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/04294
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/416,257
FILING DATE: 04-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Bialock, Donna K.
REGISTRATION NUMBER: 38,082
REFERENCE/DOCKET NUMBER: X9239
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-277-1090
TELEFAX: 317-276-3861

;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1089 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
PCT-US96-04294-1

Query Match 7.2%; Score 81; DB 5; Length 1089;
Best Local Similarity 53.7%; Pred. No. 1.9e-16;
Matches 196; Conservative 0; Mismatches 160; Indels 9; Gaps 1;

QY 6 CTGCGACCCCGGCGAGCGGCGACGTGTCATGGAAGAAATGTTCTGCGGCGTCT 65
DB 857 CTCCACACCATGCGAGGTGGCCCTGCTGATGTCAGATCAGCTCCACGAGCGTCT 798
QY 66 GGTGCATCCCGAGTGGGTGCTGTGTCAGCCGACACTGTTCCAGAACTCTACACATCG 125
DB 797 GGTCAATGAGGGGTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 738
QY 126 GCTGGGCGTGCACAGTCTTGAGGCCGACCAAGACCGAGGACGATGTTGAGGCCAG 185
DB 737 CTTGGGCGATGATACGCTGGGCGACAGAGAG-----CTCAGAGATCAAGGCTTC 687
QY 186 CTTCTCCGTAGCGGACCGACAGACAGACAGACCTTGTGCTAGACAGACCTGATCAT 245
DB 686 GAAGTCAATTCGCGACCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 627
QY 246 CAAGTTGAGCAATCCGTGTCGAGTGTACACCATCCGAGCATGCTGCTGCTGCTGCTG 305
DB 626 GAAGCTCAATAGCCAGGCGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 567
QY 306 GTGCGCTACCGCGGAGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 365
DB 566 CTGGAACCCCGTGAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 507
QY 366 TGTGA 370
DB 506 TGTGA 502

RESULT 4

PCT-US96-04294-3/C

; Sequence 3, Application PC/TUS9604294

; GENERAL INFORMATION:

; APPLICANT: Dixon, Eric P.

; APPLICANT: Johnstone, Edward M.

; APPLICANT: Little, Sheila P.

; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Eli Lilly and Company

; STREET: Lilly Corporate Center

; CITY: Indianapolis

; STATE: Indiana

; COUNTRY: United States of America

; ZIP: 46285

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US96/04294

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/416,257

; FILING DATE: 04-APR-1995

; ATTORNEY/AGENT INFORMATION:

;; NAME: Blalock, Donna K.
;; REGISTRATION NUMBER: 38,082
;; REFERENCE/DOCKET NUMBER: X9239
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 317-277-1090
;; TELEFAX: 317-276-3861
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1089 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: mRNA
PCT-US96-04294-3

Query Match 7.2%; Score 81; DB 5; Length 1089;
Best Local Similarity 53.7%; Pred. No. 1.9e-16;
Matches 196; Conservative 0; Mismatches 160; Indels 9; Gaps 1;

QY 6 CTGCGACCCCGGCGAGCGGCGACGTGTCATGGAAGAAATGTTCTGCGGCGTCT 65
DB 857 CTCCACACCATGCGAGGTGGCCCTGCTGATGTCAGATCAGCTCCACGAGCGTCT 798
QY 66 GGTGCATCCCGAGTGGGTGCTGTGTCAGCCGACACTGTTCCAGAACTCTACACATCG 125
DB 797 GGTCAATGAGGGGTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 738
QY 126 GCTGGGCGTGCACAGTCTTGAGGCCGACCAAGACCGAGGACGATGTTGAGGCCAG 185
DB 737 CTTGGGCGATGATACGCTGGGCGACAGAGAG-----CTCAGAGATCAAGGCTTC 687
QY 186 CTTCTCCGTAGCGGACCGACAGACAGACAGACCTTGTGCTAGACAGACCTGATCAT 245
DB 686 GAAGTCAATTCGCGACCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 627
QY 246 CAAGTTGAGCAATCCGTGTCGAGTGTACACCATCCGAGCATGCTGCTGCTGCTGCTG 305
DB 626 GAAGCTCAATAGCCAGGCGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 567
QY 306 GTGCGCTACCGCGGAGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 365
DB 566 CTGGAACCCCGTGAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 507
QY 366 TGTGA 370
DB 506 TGTGA 502

RESULT 5

PCT-US95-06157-5

; Sequence 5, Application PC/TUS9506157

; GENERAL INFORMATION:

; APPLICANT: Mayo Foundation for Medical Education

; APPLICANT: And Research

; APPLICANT: Hydriltech Incorporated

; APPLICANT: Tindall, Donald J.

; APPLICANT: Young, Charles Y.F.

; APPLICANT: Saeed, Mohammed S.

; TITLE OF INVENTION: Recombinant HK2 Polypeptide

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Schweigman, Lundberg & Woessner, P.A.

; STREET: 3500 IDS Center

; CITY: Minneapolis

; STATE: MN

; COUNTRY: USA

; ZIP: 55402

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent Release #1.0, Version #1.25

```

Query Match Similarity      6.8%;   Score 76.4;   DB 5;   Length 832;
Best Local Similarity      63.0%;   Pred. No. 4,8e-15;
Matches 136;   Conservative 0;   Mismatches 76;   Indels 4;   Gaps 1

0Y  311 CTACCGCGGGGAACCTCTTGCCTGTTTCAGGCTGGGCTGCTGCGGAACGATGCTGTGA 370
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   5 CCAGCATGAGGACCTGGTCTCTTCATCGCCTTGTGCTGTGGGGTGCACATGGGCGCTGC 64

0Y  371 TTGCATCCAGTCCAGACTGTGGGAGGCTGGGAGTGTAGAAGCTTTCCCAACCTGGC 430
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   65 CCCTCATCCAGTCTGGGATGTGGGAGGCTGGGAGTGTAGAAGCATTTCCCAACCTGGC 124

0Y  431 AGG----GTTTACCACTTGGGCAACTTCAGTGCAGGAAGACGCGCTGATCCTCACT 486
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   125 AGGTGGCTGTGACACTATGATGGGCACTATGTGGGGTGTCTGTGTACACCCCACT 184

0Y  487 GGGTCTCACTACTGCTCACTGCATCCACCGGACA 522
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   185 GGGTCTCACTACTGCTGCATCCATGGCTTAAGAAGATA 220

RESULT      6
US-08-358-782D-13
: Sequence 13, Application US/08358782D
: Patent No. 5674682
: GENERAL INFORMATION:
:   APPLICANT: Croce, Carlo
:   APPLICANT: Gornella, Leonard
:   APPLICANT: Mulholland, S. Grant
:   APPLICANT: Moreno, Jose
:   APPLICANT: Fischer, Rainer
:   TITLE OF INVENTION: Methods of Detecting Micrometastasis of Prostate
:   NUMBER OF SEQUENCES: 14
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
:     STREET: One Liberty Place 46th. Floor
:     CITY: Philadelphia
:     STATE: PA
:     ZIP: 19103
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patent Release #1.0, Version #1.25
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/358,782D
:     FILING DATE: 15-DEC-1994
:   CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:

```

Query Match	6.5%;	Score 72.8;	DB 2;	Length 992;
Best Local Similarity	59.9%;	Pred. No. 7.7e-14;		
Matches 139;	Conservative	0;	Mismatches 89;	Indels 4;
				Gaps 1

RESULT 7
 US-08-358-782D-14
 ; Sequence 14 Application US/08358782D
 ; Patent No. 5674682
 ; GENERAL INFORMATION:
 APPLICANT: Croce, Carlo
 APPLICANT: Gromella, Leonard
 APPLICANT: Mulholland, S. Grant
 APPLICANT: Moreno, Jose
 APPLICANT: Fischer, Rainer
 TITLE OF INVENTION: Methods of Detecting Micrometastasis of Prostate
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
 STREET: One Liberty Place 46th. Floor
 CITY: Philadelphia
 STATE: PA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/358,782D
 FILING DATE: 15-DEC-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Beardell, Lori Y.
 REGISTRATION NUMBER: 34,293
 REFERENCE/DOCKET NUMBER: TTU-1327
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-568-3100
 TELEFAX: 215-568-3439

```

? INFORMATION FOR SEQ ID NO: 14:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1462 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEetical: NO
? ANTI-SENSE: NO
?
US-08-358-782D-14

```

Query Match	6.5%	Score 72.6;	DB 2;	Length 1462;
Best Local Similarity	65.9%	Pred. No. 1.1e-13;		
Matches 122; Conservative	0;	Mismatches 59;	Indels 4;	Gaps 1

OY	410	AGAAACATTTCCCAACCCTGGCAGG----	GTTGTACCATTTTGGCAACTTCACATGCCAAGG	465
Db				
	119	AGAAGACATTTCCCAACCCTGGCAGGCTGTG	TGGGCTCTCGTGCGGCAAGGGCAGTGTCCGGCG	178
Db				
OY	466	ACGCTTCCTGCCTACCTCTACTCTGGTGTCT	CCTACTCTACTGTGATCACCCGGAACTGTG	525
OY				
OY	179	GTTGTTCTGGTGTACACCCCAAGTGGGTCT	CTCCTACAGCTGCCCACTGCATCAGGAACA	238
Db				
OY	526	TGATC	530	
Db	239	TGATC	243	

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match 6.48; Score 72; DB 3; Length 200;
Best Local Similarity 66.98; Pred. No. 5.4e-14;
Matches 119; Conservative 0; Mismatches 55; Indels 4; Gaps 1.

[illegible]

```

RESULT      9
PCT-US95-06157-9
: Sequence 9, Application PC/TUS9506157
: GENERAL INFORMATION:
:   APPLICANT: Mayo Foundation for Medical Education
:   APPLICANT: and Research
:   APPLICANT: Hybritech Incorporated
:   APPLICANT: Tindall, Donald J.F.
:   APPLICANT: Young, Charles Y.F.
:   APPLICANT: Saedi, Mohammed S.
:   TITLE OF INVENTION: Recombinant HK2 Polypeptide
:   NUMBER OF SEQUENCES: 18
:   CORRESPONDENCE ADDRESS:
:   ADDRESSEE: Schegman, Lundberg & Woessner, P.A.
:   STREET: 3500 IDS Center
:   CITY: Minneapolis
:   STATE: MN
:   COUNTRY: USA
:   ZIP: 55402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/06157
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
:   NAME: Raasch, Kevin W.
:   REGISTRATION NUMBER: 35,561
:   REFERENCE/DOCKET NUMBER: 150.148W01
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: 612-339-0331
:   TELEFAX: 612-339-3061
:   INFORMATION FOR SEQ ID NO: 9:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 766 base pairs
:     TYPE: nucleic acid
:     STRANDEDNESS: double
:     TOPOLOGY: linear
:     MOLECULE TYPE: cDNA
:     FEATURE:
:       NAME/KEY: CDS
:       LOCATION: 1..732
: PCT-US95-06157-9

```

```
Query Match      6 48; Score 71.6; DB 5; Length 766;  
Best Local Similarity 71.7%; Pred. No. 1.6e-13;  
Matches 109; Conservative 0; Mismatches 39; Indels 4; Gaps 1.  
  
QY    375 CATCCAGTCCCAAGACTGTGGGAGCGTGGAATGTGAGAAAGCTTTCCCAACCTTCGCAGG- 434  
|||||
```

Db 9 CATTGATCTCGGATTTGTGGAGGCTGGAGATGTGAGAACATTTCCCAACCTTGGCAGCT 68
QY 434 ---GTTGTACCAATTTGGGCACTTCCAGTGCAGACGTCCTGCTCATCTTCCAGCGGT 490
Db 69 GCGTGTGTACGATCATGTGGGACACATGTGGGAGTGTCTGTGTGATGACCCCGCAGTGGT 128
QY 491 GCTCACTACTGCTCATCATCACCAGGACA 522
Db 129 GCTCACAGCTGCCCATTTGCTTAAGAAAGATA 160

RESULT 10
US-08-844-024-1
Sequence 1, Application US/08844024
Patent No. 5840494
GENERAL INFORMATION:
APPLICANT: Katz, Aaron E., et al.
TITLE OF INVENTION: A Method For Molecular Staging Of
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/844,024
FILING DATE: 18-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/229,391
FILING DATE: 15-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 43677/JPW/TEP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422533 COOP UI
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1729 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 378..1088
US-08-844-024-1

Query Match 6.1%; Score 68; DB 4; Length 1729;
Best Local Similarity 66.1%; Pred. No. 3.7e-12;
Matches 115; Conservative 0; Mismatches 55; Indels 4; Gaps 1;

QY 361 GATGCTGTGATTCATCCAGTCCAGACTGTGAGAGCTGGAGATGTGAGAGCTTTC 420
Db 351 GGTGTGACACCCCTGATCTGTGATTTGGAGGCTGGAGATGTGAGAGATTC 410
QY 421 CAACCTGGCAGG---GTTGTACCAATTTGGGCACTTCCAGTGCAGACGTCCTGCTG 476
Db 411 CACCTGGAGGCTGTGTGAGCTCTCGTGGCAGGACAGTCTGCGGCGGTGTTCTGCTG 470

QY 477 CATCTCACTGGAGTGTCTCATCTGCTCATCTGATCATCCAGGACACACTGTGATC 530
Db 471 CACCCCCAGTGGTCTCTCACTGACTTACCACATGATCAGGAACAAAGCTGATC 524

RESULT 11
PCT-US95-06157-7
Sequence 7, Application PC/TUS9506157
GENERAL INFORMATION:
APPLICANT: Mayo Foundation for Medical Education
APPLICANT: and Research
APPLICANT: Hybritech Incorporated
APPLICANT: Tindall, Donald J.
APPLICANT: Young, Charles Y.F.
APPLICANT: Saeed, Mohammed S.
TITLE OF INVENTION: Recombinant HK2 Polypeptide
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.
STREET: 3500 IDS Center
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06157
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Raasch, Kevin W.
REGISTRATION NUMBER: 35,561
REFERENCE/DOCKET NUMBER: 150.148W01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-339-0331
TELEFAX: 612-339-3061
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 760 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 7..720
PCT-US95-06157-7

Query Match 5.7%; Score 63.6; DB 5; Length 760;
Best Local Similarity 58.6%; Pred. No. 5.9e-11;
Matches 130; Conservative 0; Mismatches 89; Indels 3; Gaps 1;

QY 3 GCATCTGCAGCCCTGGCAGGCGCACTGTGATGAAAAAGATTGTCTGCTCGGGCGT 62
Db 36 GCATTCGCCAACCTCGGACAGTGTGTGATGATGATGATGATGATGATGATGATGATG 95
QY 63 CCTGTGTGATCCGCGAGTGGTGTCTGTGACGCCGACACACTTTTCCAGAACTCTTACACAT 122
Db 96 CCGTGTGACCCCGCCAGTGGTGTCTCACACCTGCCATTTGCTTAAGAAATATAGCCAGGT 155
QY 123 CGGGCTGGGCTGCACAGTCTTGAGGCGCAGCAAGAGCCAGGAGGAGATGTGGAGGC 182
Db 156 CTGGCTGGTTCGGACACACTGTTTGTAGCTTAAGACACAG---GCCAGAGGCTCCCTGT 212
QY 183 CAGCTCTTCGTAAGGACCCAGAGTACACAGAGACCTTGT 224
Db 213 CAGCCACAGCTTCCACACCCGCTCTACAAATATGAGCCTTCT 254

OY	283	CGGACATCACCAT-----TGCCTTGAGAGCCCTCAGGGGGAACCTGTGCTTCGT	336
Db	419	CACAGGTCACAGTCGTCCCCCTGCTCGGAGACTTCCCCTCGGGGATGCCGTGCTGCTC	478
OY	337	TCTGGCTGGGGTCTCTGTCGCCAAGATCGCTGTGATTGCATCACTCCGAACTGTGGGA	386
Dd	479	ACTGCTGTGGGCGACAGCTGGACAATAATGTGGTGTGGGACACGGCGGAGGCCCA	538
OY	397	GAGTTGGGA	404
Dd	539	GATGGCA	546

RESULT 14
US-08-845-998-5
; Sequence 5, Application US/08845998
; Patent No. 5879892

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```

?      COUNTRY:  US
?      ZIP:  02210
?
?      COMPUTER READABLE FORM:
?
?      MEDIUM TYPE:  Floppy disk
?
?      COMPUTER:  IBM PC compatible
?
?      OPERATING SYSTEM:  PC-DOS/MS-DOS
?
?      SOFTWARE:  Patent In Release #1.0, Version #1.25
?
?      CURRENT APPLICATION DATA:
?
?      APPLICATION NUMBER:  US/08/845,998
?
?      FILING DATE:
?
```

1 CLASSIFICATION: 435
2 ATTORNEY/AGENT INFORMATION:
3 NAME: Van Amsterdam, John R.
4 REGISTRATION NUMBER: 40,212
5 REFERENCE/DOCKET NUMBER: L0461/7008
6 TELECOMMUNICATION INFORMATION:
7 TELEPHONE: (617)720-3500
8 TELEFAX: (617)720-2441

```

? INFORMATION FOR SEQ ID NO: 5
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2218 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? IMMEDIATE SOURCE:
? CLONE: NVB352/3
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 8..577
US-08-845-998-5
```

Query Match	5.1%	Score 56.8	DB 4	Length 2218
Best Local Similarity	51.4%	Pred. No. 1.7e-08		
Matches 169; Conservative	0	Mismatches 167	Indels 12	Gaps 2

```
QY      49  TTCTGTCGGGCGCTCCGTGTCATCCGGAGTGGGTGCTGTACGCCGCACACTGTT----- 105
        ||||| |||| | | | | | | | | | | | | | | | | | | | | | | | |
Db     179  TTCTGGGGGGGCTCCTTCATCACCCCCAGTAGGGTGCTAACCGCGGGGCACTGCGGTGA 238
```

QY	105	-CCAGAACCTCTACACACATCGGGCTGGGCTCTGCACAGCTTGATGGCCGACCAAGAGCCA	162
Db	239	CCGAGCATCAAGAGTCTGGCCCGCCCTCAGGCTGCACCTGGGAGACGCACTCTACTAC	298
QY	163	GGAGACCAAGATGGTGGAGGCCAGCCTCTCCGTACGGCACCCAGATACACAGACCTTG	222
Db	299	CAGGATCCAGCTGCTGGCCGGTCCAGAGGATATCTGTGACCCACAGTTCTTACATCAATCAG	358
QY	223	CTCGCTAACGACCTCATGCTCATCAAGTTGAGAGATCCGCTGTCCGAGTCTGACACCATC	282
Db	359	ACCGGGGGGACATCGCCCTGCTGAGACTGGAGAGCCCGTGAAACATCTCCAGCCACATC	418
QY	283	CGGAGCATCAECAT-----TGCCTTGCATAGTCCCTACCCCGGGGGAATCTTGCCCTCGTT	336
Db	419	CACACGGTCAAGCTGACCCCTCCCTCCGGAACCTTCCCTCCCGGGGGAATGCGCTGGGTC	478
QY	337	TCTGGCTGGGGCTCTGGCTGGGGCAAGATGCTGTATTTCCTATCCAGTCCAGATCTGGGA	396
Db	479	ACTGGCTGGGGCGACGAGGACATATATGTGGGTGGGGGACAGCGCGGAGCGGGGCCA	538
QY	397	GGCTGGGA	404
Db	539	GGTGGCA	546

RESULT 15
US-08-684-862-11
; Sequence 11, Application US/08684862
; Patent No. 5759541

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```

1 STATE: D.C.
2 COUNTRY: USA
3 ZIP: 20036
4
5 COMPUTER READABLE FORM:
6 MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
7
8 COMPUTER: IBM AT-compatible, 80286 processor
9 OPERATING SYSTEM: MS-DOS version 5.0
10 SOFTWARE: Wordperfect version 5.1

```

```

1      CURRENT APPLICATION DATE:
2      APPLICATION NUMBER:  US/08/684,862
3
4      FILING DATE:
5
6      CLASSIFICATION:  435
7
8      PRIOR APPLICATION DATA:
9      APPLICATION NUMBER:  US/08/361,705
10
11     FILING DATE:
12
13     PRIOR APPLICATION DATA:
14     APPLICATION NUMBER:  07/966,040
15
16     FILING DATE:  30-DEC-1992
17
18     APPLICATION NUMBER:  PCT/EP91/01361
19
20     FILING DATE:  19-JUL-1991
21
22     INFORMATION FOR SEQ ID NO:  11:
23
24     SEQUENCE CHARACTERISTICS:
25
26     LENGTH:  957 base pairs

```

```

1      TOPOLOGY: linear
2
3      MOLECULE TYPE: cDNA to mRNA
4
5      ORIGINAL SOURCE:
6
7      ORGANISM: Agkistrodon rhodostoma
8
9      FEATURE:
10
11         LOCATION: 210 to 911
12
13         OTHER INFORMATION: the coding region shown in (2)(1x)(B)
14
15         OTHER INFORMATION: codes for the protein of SEQ ID NO: 6
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 12:05:28 ; Search time 1811.29 Seconds
(without alignments)
1218.615 Million cell updates/sec

Title: US-09-030-606-177
Perfect score: 1119
Sequence: 1 GCGCACTCCAGCCCTGCA.....ATCTATAAAAAAAAAAAAA 1119

Scoring table: IDENTITY_NUC
2546578 seqs, 98626752 residues

Database : EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: em_est20:*
21: em_est21:*
22: em_est22:*
23: em_est23:*
24: em_est24:*
25: em_est25:*
26: em_est26:*
27: em_est27:*
28: em_est28:*
29: em_est29:*
30: em_est30:*
31: em_est31:*
32: em_est32:*
33: em_est33:*
34: em_est34:*
35: em_est35:*
36: em_est36:*
37: em_est37:*
38: em_est38:*
39: em_est39:*
40: em_est40:*
41: em_est41:*
42: em_est42:*
43: em_est43:*
44: em_est44:*
45: em_est45:*
46: em_est46:*
47: em_est47:*
48: em_est48:*
49: em_est49:*
50: em_est50:*
51: em_est51:*
52: em_est52:*
53: em_est53:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	361	32.3	777	48	AI557281	AI557281 PT2.1-15-
2	223.6	20.0	722	48	AI557025	AI557025 PT2.1-10-
3	214.2	19.1	415	35	AA551449	AA551449 nj55e05.s
4	205.8	18.4	569	50	AI686689	AI686689 tuj35g11.x
5	170	15.2	216	32	AA336074	AA336074 EST40886
6	99.6	8.9	479	46	AI436093	AI436093 th92b12.x
7	99.6	8.9	475	46	AI452907	AI452907 t46g04.x
8	92	8.2	245	31	AA295609	AA295609 EST100780
9	91.6	8.2	264	32	AA366980	AA366980 EST77998
10	82.2	7.3	429	35	AA594076	AA594076 n31a12.s
11	81.4	7.3	407	38	AA776018	AA776018 aeb3g11.s
12	81	7.2	523	35	AA542894	AA542894 nj55b01.s
13	76.2	6.8	486	35	AA535837	AA535837 nj79g05.s
14	76.2	6.8	572	35	AA579026	AA579026 nj34f02.s
15	74.8	6.7	253	20	T29510	T29510 EST82448 Hu
16	74	6.6	546	33	AA388907	AA388907 mp16b03.r
17	72.6	6.5	336	34	AA506459	AA506459 nh46a08.s
18	72.6	6.5	548	35	AA578976	AA578976 nf26g03.s
19	72.6	6.5	515	35	AA593245	AA593245 n07e10.s
20	72.6	6.5	902	47	AI525832	AI525832 PT1.3-06-
21	72.6	6.5	732	48	AI547309	AI547309 PNO01.AH
22	71.6	6.4	470	47	AI525128	AI525128 p10rma-7
23	71	6.3	610	48	AI557591	AI557591 pt2.1-1.f
24	70	6.3	371	34	AA503943	AA503943 nh38d05.s
25	70	6.3	484	36	AA603329	AA603329 np15b08.s
26	70	6.3	523	36	AA639901	AA639901 np08f03.s
27	70	6.3	737	48	AI547285	AI547285 PNO01.AH
28	70	6.3	454	50	AI685510	AI685510 tuj3b10.x
29	69.8	6.2	703	48	AI546857	AI546857 PN2.1-07-
30	69.4	6.2	457	45	AI385433	AI385433 m185b05.y
31	69.4	6.2	595	48	AI557389	AI557389 PT2.1-6-G
32	69	6.2	619	26	W58737	W58737 zd33e07.r.1
33	69	6.2	435	29	AA177005	AA177005 nc03b02.s
34	68.8	6.1	432	38	AA778667	AA778667 af87e13.s
35	67.8	6.1	340	27	AA038537	AA038537 m185b05.r
36	67.2	6.0	519	42	AI098337	AI098337 v984b03.r
37	67.2	6.0	760	42	AI127299	AI127299 qb74f07.x
38	66.8	6.0	453	34	AA533652	AA533652 nj72g08.s
39	65.6	5.9	616	48	AI546898	AI546898 PN2.1-09-
40	65.4	5.8	457	42	AI078417	AI078417 oz05e02.x
41	65	5.8	624	47	AI525001	AI525001 p10rma-4
42	64.2	5.7	397	35	AA573604	AA573604 nj42d12.s
43	64	5.7	472	28	AA062294	AA062294 m170a06.r
44	64	5.7	482	28	AA066464	AA066464 m173c10.r
45	64	5.7	568	28	AA068804	AA068804 mm62g08.r

ALIGNMENTS

RESULT 1
LOCUS AI557281 777 bp mRNA
DEFINITION PT2.1-15_G12.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION AI557281
NID 94489644
VERSION AI557281.1 GI:4489644
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS 1 (bases 1 to 777)
Huang, G.M., Ng, W., Farkas, J., Chen, L., Liang, H.A., Gordon, D., Jun Yu, J., and Hood, L.
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138767.

CONTACT: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com

FEATURES
source
1..777
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="960H11: 6; 6p21.31-6p22.1; 21q"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional cDNA library was constructed using Lambda ZP II kit (Stratagene). mRNA was extracted from a frozen prostate tumor tissue (Mayo Clinics)."

BASE COUNT 162 a 205 c 227 g 157 t 26 others
ORIGIN

Query Match 32.3%; Score 361; DB 48; Length 777;
Best Local Similarity 100.0%; Pred. No. 6.9e-96;
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCACTCGAGCCCTGGCAGGCGGCACTGTCATGGAACGAATGTTCTGCTGGGC 60
DB 14 GCGCACTCGAGCCCTGGCAGGCGGCACTGTCATGGAACGAATGTTCTGCTGGGC 73
QY 61 GTCTCTGTGATCCGAGTGGTGTGTACGCCGACACTGTTCCAGAACTCTTACACC 120
DB 74 GTCTCTGTGATCCGAGTGGTGTGTACGCCGACACTGTTCCAGAACTCTTACACC 133
QY 121 ATCGGCTGGGCTTGCACAGTCTTGAAGCGGCAAGAGCCAGGAGCCGATGTTGAG 180
DB 134 ATCGGCTGGGCTTGCACAGTCTTGAAGCGGCAAGAGCCAGGAGCCGATGTTGAG 193
QY 181 GCCAGCTCTCGCTACGAGCCAGAGATACAGACCCCTTGTCTGCTAAGCACTCATG 240
DB 194 GCCAGCTCTCGCTACGAGCCAGAGATACAGACCCCTTGTCTGCTAAGCACTCATG 253
QY 241 CTGATCAAGTGGAGCAATCCGTGTCCGAGTCTGACACCACTCCGAGCATCAGATTGCT 300
DB 254 CTGATCAAGTGGAGCAATCCGTGTCCGAGTCTGACACCACTCCGAGCATCAGATTGCT 313
QY 301 TCGCAGTGGCCTTACGCGGGGGAACCTTGTGCTTGTGCGGGGCTGCGTGGGGAAC 360
DB 314 TCGCAGTGGCCTTACGCGGGGGAACCTTGTGCTTGTGCGGGGCTGCGTGGGGAAC 373
QY 361 G 361
DB 374 G 374

RESULT 2
LOCUS A1557025 722 bp MRNA EST 23-MAR-1999
DEFINITION PT2.1.10_F05.r tumor2 Homo sapiens cDNA 3', mRNA, sequence.
ACCESSION A1557025
NID 94489388
VERSION A1557025.1 GI:4489388

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS 1 (bases 1 to 722)
Huang, G.M., Ng, W., Farkas, J., Chen, L., Liang, H.A., Gordon, D., Jun Yu, J., and Hood, L.
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138511.

CONTACT: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com

FEATURES
source
1..722
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="15; 21q"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional cDNA library was constructed using Lambda ZP II kit (Stratagene). mRNA was extracted from a frozen prostate tumor tissue (Mayo Clinics)."

BASE COUNT 146 a 156 c 195 g 176 t 49 others
ORIGIN

Query Match 20.0%; Score 223.6; DB 48; Length 722;
Best Local Similarity 95.6%; Pred. No. 2.1e-55;
Matches 237; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 109 AACTCTTACACCATGGGCTGGGCTTGAAGCGGCAAGAGCCAGGAGGC 168
DB 115 AACTCTTACACCATGGGCTGGGCTTGAAGCGGCAAGAGCCAGGAGGC 174
QY 169 CAGATGTGGAGCCAGCCTTCGCTACGAGCCAGAGATACAGACCCCTTGTCTGCT 228
DB 175 CAGATGTGGAGCCAGCCTTCGCTACGAGCCAGAGATACAGACCCCTTGTCTGCT 234
QY 229 AACGACCTTATCTATCAATGTTGACGATCCGTGTCCGAGTCTGACACCACTCCGAGC 288
DB 235 AACGACCTTATCTATCAATGTTGACGATCCGTGTCCGAGTCTGACACCACTCCGAGC 294
QY 289 ATCAGCATGCTTCCGAGAGCCCTACCGGCGGGAACCTTGTCC-TCGTTTCGCTGGG 347
DB 295 ATCAGCATGCTTCCGAGAGCCCTACCGGCGGGAACCTTGTCCGCTGCTGGG 354
QY 348 TCTGCTGG 355
DB 355 TCTGCTGG 362

RESULT 3
LOCUS AA551449 415 bp MRNA EST 05-SEP-1997
DEFINITION n155e05.s1 NCI CGAP pr9 Homo sapiens cDNA clone IMAGE:996416 similar to SW:K1KA_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR ;, mRNA sequence.
ACCESSION AA551449
NID 92321701
VERSION AA551449.1 GI:2321701
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

VERSION AA336074.1 GI:1988560
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 216)
Adams,M.D., Kerlavage,A.R., Fleischman,R.D., Fuldner,R.A.,
Balt,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-ai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geisbagen,N.S.,
Glodde,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
Keller,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudak,D.M., Shiley,R.,
Small,K.V., Spriggs,T.A., Utechtack,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepede,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,M.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wel,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Kosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M., and Venter,J.C.

TITLE Initial assessment of human gene diversity and expression patterns
JOURNAL based upon 83 million nucleotides of cDNA sequence
MEDLINE Nature 377 (6547 Suppl), 3-174 (1995)
COMMENT On Apr 14, 1993 this sequence version replaced gi:692773.

CONTACT: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3016699056
Fax: 3016699423
Email: arkerlavage@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES
Location/Qualifiers
1..216
/organism="Homo sapiens"
/db_xref="ATCC (inhost):138014"
/db_xref="taxon:9606"
/clone_lib="Endometrial tumor"
/sex="Female"
/dev_stage="adult"
/note="Organ: endometrium; Vector: pluescript SK-";
Site_1: EcoRI, Site_2: XhoI
BASE COUNT 45 a 65 c 63 g 38 t 5 others
ORIGIN

Query Match 15.2%; Score 170; DB 32; Length 216;
Best Local Similarity 97.1%; Pred. No. 8.1e-40;
Matches 170; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
DB 2 CGCACTCGACCGCTGGAGCGGCGCATGTCATGGAAGCAATGTTGCTCGGGCG 61
DB 42 CGCACTCGACCGCTGGAGCGGCGCATGTCATGGAAGCAATGTTGCTCGGGCG 101
DB 62 TCCGAGTCCGACGAGGCGGCTGTCAGCGGCGCATGTTTCCAGAACCTCAGACCA 121
DB 102 TCCGAGTCCGACGAGGCGGCTGTCAGCGGCGCATGTTTCCAGAACCTCAGACCA 161
DB 122 TCGGGCTGGGCTGACAGTCTTGAGCGCGGACCAAGAGCCAGGAGGAGATGTG 176
DB 162 TCGGGCTGGGCTGACAGTCTTGAGCGCGGACCAAGAGCCAGGAGGAGATGTG 216
RESULT 6

AI436093/c
LOCUS AI436093 479 bp mRNA EST 30-MAR-1999
DEFINITION tb92b12.x1 Soares NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
IMAGE:2126111 3' similar to TR:002710 002710 GAG POLYPROTEIN.;
mRNA sequence.
AI436093
94307772
VERSION AI436093.1 GI:4307772
KEYWORDS EST.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 479)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ccgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

TITLE Tumor Gene Index
JOURNAL On May 18, 1998 this sequence version replaced gi:3137463.
COMMENT

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 677 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 444.

FEATURES
Location/Qualifiers
1..479
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAG:2126111"
/clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NBHP pool 1:
309384-310919, 323208-325895 Soares NBHP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares NBHP pool 1:
758280-760583, 772104-774407 Soares NBHP pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHP
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 143 a 85 c 127 g 124 t
ORIGIN

Query Match 8.9%; Score 99.6; DB 46; Length 479;
Best Local Similarity 62.5%; Pred. No. 6.4e-19;
Matches 240; Conservative 0; Mismatches 134; Indels 10; Gaps 5;
DB 557 AGTCAGACTATCATGATTACTGTGTGACGTGCTGTATTTGATTAACATGCC--- 613
DB 458 AGACAGACTATTCATGTTGCTGCTTCCCTGCTGTGTCATTTAGAGAGTACACCCACC 399
DB 613 --GATGTTAGGTGAATTTAGCGTCACTTGCGCTCAACCACTTTGGTATCCAGTTATCT 670
DB 398 TTGGCTTTGATGTTGATGCTGTCACTTGCGCACTGCCACTTCAGAGATCAATTTATTC 339
DB 671 CACGATTTGATGTTGCTGCTGTCAGTGCAGCATTCGAC--ATAATTCTGACCGACA 729
DB 338 CATTGATTTAATTTTGTGATGAGTGTGATGCTCCACGTTTGAATGATGATCAATCC 279
DB 730 GAGGTGAGGATCATATAGCTTTCAAGATGCTGTACTCCCTCAATTCATTTCT 789
II III II I IIIIIIIII IIII III IIIIIIIIIII IIII

8.98; Score 99.6; DB 46; Length 475;

On Nov 29 1993 this sequence version replaced gl:635027
Other_ESTS: TH0167057
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
5712 Medical Center Drive, Rockville, MD 20850 USA

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 523)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
On Sep 12, 1996 this sequence version replaced gi:1405040.
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Insert Length: 1057 Std Error: 0.00
Seq primer: -21ml3 forward (Amersham)
High quality sequence stop: 462.
FEATURES
Location/Qualifiers
1..523
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:980713"
/clone_lib="NCI_CGAP_Ov2"
/sex="female"
/tissue_type="ovary"
/lab_host="DH10B"
/note="Vector: PAMP10; mRNA made from invasive ovarian
tumor; CDNA made by oligo-dT priming. Non-directionally
cloned. Size-selected on agarose gel, average insert size
600 bp. Reference: Kitzman et al. (1996) Cancer Research
56:5380-5383."
BASE COUNT 112 a 171 c 138 g 102 t
ORIGIN
Query Match 7.2%; Score 81; DB 35; Length 523;
Best Local Similarity 53.7%; Pred. No. 2e-13; Indels 9; Gaps 1;
Matches 196; Conservative 0; Mismatches 160;
QY 6 CTCGACCCCTGGCAGCGGCGACTGTCATGAAACGAATTGCTGCGGCGCTCT 65
DB 121 CTCGACCCCTGGCAGCGGCGACTGTCATGAAACGAATTGCTGCGGCGCTCT 180
QY 66 GGTGATCCGCAATGGTGTCTGTCAGCCGACACTGTTCCAGAACTCTTACACATCGG 125
DB 181 GGTCAATGACCGGTGGTGTCTCACTGCCGCCACTGCAGATGATGATACACCGTCA 240
QY 126 GGTGGGCTGTCACAGTCTTGAAGCCGACAGAGCCAGAGCCAGATGTTGAGGCGCAG 185
DB 241 CTTGGGAGTGAATGCTGGGCGACAGAGAG-----CTCAGAGATCAAGGCTTC 291
QY 186 CCTCTCCGTAGGCGACCCAGACAGACAGCCCTTGTCTGCTAACGACTCATGTCAT 245
DB 292 GAGTCAATTCGGACACCCCGGCTACTTCACACAGACCAAGTATATACCTCATGCTCT 351
QY 246 CAAGTTGAGCAATCCGTGTCAGTCTGACACCAATCCGAGAGATCGATTCGCTTGC 305
DB 352 GAAGCTCAATAGCCAGGCGCTGTCATCATGTCATGTAAGTAAGTACGAGCTCCCG 411
QY 306 GTGCGCTACCGGGGGGGAACCTTGTCTGCTGGGCGTCTGCTGGCGAAGCATGC 365
DB 412 CTCGGAACCCCTGGACCACTGTACTGTCTCGGCTGGGCGACTACCAAGAGCCCGA 471

QY 366 TGTGA 370
DB 472 TGTGA 476
RESULT 13
AA535837
LOCUS
DEFINITION
AA535837 486 bp mRNA EST 21-AUG-1997
n179g05.s1 NCI_CGAP_P10 Homo sapiens cDNA clone IMAGE:988744
similar to gb:539325 GLANDULAR KALLIKREIN 2 PRECURSOR (HUMAN);,
mRNA sequence.
ACCESSION
AA535837
NID 92280090
VERSION AA535837.1 GI:2280090
KEYWORDS
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 486)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
On Apr 14, 1993 this sequence version replaced gi:315330.
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuquib,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Insert Length: 1091 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 385.
FEATURES
Location/Qualifiers
1..486
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:988744"
/clone_lib="NCI_CGAP_P10"
/sex="male"
/tissue_type="invasive prostate tumor"
/lab_host="DH10B"
/note="Organ: prostate; Vector: PAMP10; mRNA made from
invasive prostate tumor; CDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 600 bp. Library made by D. Kitzman,
NIH."
BASE COUNT 96 a 142 c 145 g 103 t
ORIGIN
Query Match 6.8%; Score 76.2; DB 35; Length 486;
Best Local Similarity 66.8%; Pred. No. 5.1e-12;
Matches 125; Conservative 0; Mismatches 58; Indels 4; Gaps 1;
QY 340 GGTGGGGTCTGTGCGAAGCATGCTGTGATGTCATCCAGTCCAGACTGTGGAGGC 399
DB 58 GCTTCTCTGTGGGGTGCAGTGTGCTGCCCTGCATCATGCTCGATTTGGGAGGC 117
QY 400 TTGGAGTGTGAGAAAGTTCCCAACCTCGCAGG-----GTGACCATTTGGCAACTTC 455
DB 118 TTGGAGTGTGAGAAAGTTCCCAACCTCGCAGGTTGCTGTGACGTATGATGGAGCA 177
QY 456 CAGTGAAGAGCTCTGCTGATCTCTACCTGGTGTCTCACTGCTCATGCTCATGCTACCC 515

```

Db      178 CACGTG6GGGGTCTCTGTGCATGCCACCAGTG6G9TCACAGACTGCCCATTTGCCTAAG 237
OY      516 CGGAACA 522
        ||| |
Db      238 AAGAATA 244

RESULT 14
LOCUS   AA579026          572 bp      mRNA           EST       12-SEP-1997
DEFINITION n134f02.s1 NCI_CGAP_Prl Homo sapiens cDNA clone IMAGE:915675
            similar to gb:S39329 GIANDIAR KALLIKREIN 2 PRECURSOR
            (HUMAN); contains Alu repetitive element,, mRNA sequence.
ACCESSION AA579026
NID       92357210
VERSION   AA579026.1 GI:2357210
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 572)
AUTHORS  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
          Tumor Gene Index
          Unpublished (1997)
COMMENT   On Sep 12, 1996 this sequence version replaced gi:1407214.

Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuquai,
M.D., Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Kitzman, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLW at:
www-bio.lnl.gov/bdrip/image/image.html

Insert Length: 921 Std Error: 0.00
Seg primer: -40m3 fwd. ET from Amersham
High quality sequence stop: 159.
Location/Qualifiers
1..572
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:915675"
/clone_1lb="NCI_CGAP_Prl"
/dev_stage="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/vector="pAMP10"
/site_1="NotI; site_2="EcoRI; 1st
strand cDNA was primed with oligo(dT)/17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected, histologically normal
prostate epithelial cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into pAMP10 by the ligation method
(Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Kitzman."

BASE COUNT 118 a 153 c 181 g 120 t
ORIGIN
Query Match 6.8%; Score 76.2; DB 35; Length 572;
Best Local Similarity 66.8%; Pred.No.5.5e-12;
Matches 125; Conservative 0; NoMatches 58; Indels 4; Gaps 1;
340 GGCTGGGGCTCTCTGCGCAACATCTGTGATTCGCATTCATCCCAACACTGTGGAGGC 399
|||||||

```

Db	Accession	Source	Organism	Reference	Authors
Db	16	GCATTGCTGTGGGGGTGCTACTGTGTGGCGTGGCCCTCATGCATCTCGATTGTGGGAGGC	75		
Qy	400	TGGGAGTGTGGAAGCATTTTCCCAACCCCTGGCAGG---	455		
Db	76	TGGGAGTGTGGAAGCATTTTCCCAACCCCTGGCAGTGGCTGTGATGAGTATGATGAGGCA	135		
Qy	456	CAGTGCAGGACGCGCCGTGCATCCCTGACATGGCTGCTCATACGCTCAGTCATCAGC	515		
Db	136	CACGTGTGGGGGTGTCTGTGTGCACCCCAAGTGGGTCTCATACAGTGTCCATGCTCTAAG	195		
Qy	516	CGGACA	522		
Db	196	AAGATA	202		
RESULT	15				
LOCUS	T29510	253 bp	MRNA	EST	06-SEP-1995
DEFINITION	EST82448	Human Prostate gland Homo sapiens	CDNA 5' end similar to		
ACCESSION	T29510	Kallikrein 1, renal/Pancreas/salivary (HT:111),	MRNA sequence.		
MD	9611608				
VERSION	T29510.1	GI:611608			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Primates; Catarrhini; Homiidae; Homo.				
	1 (bases 1 to 253)				
	Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,				
	Balt,C.J., Lee,N., Kirkness,E.F., Weinstein,K.G., Gocayne,J.D.,				
	White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chui,M.-W.,				
	Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,				
	Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S.M.,				
	Glocke,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkley,P.S.,				
	Kelley,J.M., Klimk,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,				
	Merlino,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T.,				
	Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,				
	Saunder,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,				
	Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A.,				
	Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrite,A.,				
	Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,				
	Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,				
	Weissner,P., S., Olsen,H., Raymond,L., Wei,Y.-F., Wing,J., Xu,C.,				
	Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,				
	Hasseltine,W.A., Fields,C., Fraser,C.M., and Venter,J.C.				
	Initial Assessment of Human Gene Diversity and Expression Patterns				
	Based Upon 83 Million Basepairs of cDNA Sequence				
	Nature 377, 3-174 (1995)				
	96026280				
JOURNAL	COMMENT				
MEDLINE					
CONTACT:	Venter, J.C.				
	The Institute for Genomic Research				
	932 Clopper Rd, Gaithersburg, MD 20878				
	Tel: 3018699036				
	Fax: 3018699423				
	Email: tdbinfo@tdb.tigr.org				
	For clone availability, additional sequence and expression				
	information related to this EST, please contact the TIGR Database				
	(tdbinfo@tdb.tigr.org)				
	Seq primer: M13 Reverse.				
FEATURES	source				
	Location/Qualifiers				
	1..253				
	/organism="Homo sapiens"				
	/db_xref="ATCC (labost):106507"				
	/db_xref="taxon:9606"				
	/clone_lib="Human Prostate gland"				
	/note="Organ: prostate gland"				
BASE COUNT	42 a	68 c	81 g	59 t	3 others
ORIGIN					

Best Local Similarity 65.8%; Pred. No. 1e-11;
Matches 123; Conservative 0; Mismatches 60; Indels 4; Gaps 1;
QY 340 GGCTGGGGTCTGCTGGGAGATGCTGATGCGATCCAGTCCAGACTGGGAGGC 399
Db 39 GCCTTGCTCTGTGGGTGACACTGGTCCCTCTCATCCAGTCTCGATTGGGAGGC 98
QY 400 TGGAGTGTGAGAGCTTTCCCAACCTGGCAGG---GTTGTACCATTTGGCACTTC 455
Db 99 TGGAGTGTGAGAGCATTCCTCAACCTGGCAGTGGCTGTACAGTCAATGATGGCA 158
QY 456 CAGTGCAGAGAGCTCCTGCTGATCCTCACTGGGTGCTCACTGCTCACTGCATCACC 515
Db 159 CACTGTGGGGTNTCTGTGTGACACCCCACTGGGTGCTCACTGCTCACTGCTCAAG 218
QY 516 CGGACA 522
Db 219 AAGATTA 225

Search completed: September 25, 1999, 12:05:32
Job time: 8136 sec


```

47 SGPHELLEASPGIULEUAlaserleuValProthrcysasnalalsetse 63
92 CCMAAAAYWATTAATTGACCATTAAT 65
63 rarglylsleuasplysleuthrvalleu 72

```

seq_name: A_Geneseq_36:W34260

seq_documentation_block:

```

ID W34260 standard; Protein: 892 AA.
AC W34260;
DT 11-MAY-1998 (first entry)
DE Rat ob receptor isoform c'.
KW Ob receptor isoform c'; obesity; leptin; ligand; rat.
OS Rattus sp.
PN MO9742340-A1.
PD 13-NOV-1997.
PF 02-MAY-1997; U07521.
PR 24-MAY-1996; GB-010995.
PR 06-MAY-1996; US-016899.
PA (MERL) MERCK & CO INC.
PI Caskey CT, Hess JW, Hey P, Phillips MS;
DR WPI: 97-556993/51.
PT New isoform(s) of the Ob (leptin) receptor - used for identifying
PT specific binding ligands, potentially useful for study, prevention
PT and treatment of obesity
PS Disclosure: Page -: 34pp; English.
CC This protein comprises isoform c' of the rat ob receptor. The ob
CC receptor has numerous isoforms resulting from alternative splicing;
CC 3 novel isoforms, designated f, g and c' are disclosed (see
CC W3458-60). Isoform c' differs from the 1162-residue wild-type
CC protein (see W34257) in that after lys-889 there are only 3 amino
CC acids. The c' isoform can be expressed in host cells, particularly
CC E. coli, yeast or mammalian cells. It is used to identify specific
CC binding ligands. Agonists, antagonists and ligand mimetics can be
CC identified that are potentially useful in the study, prevention and
CC treatment of obesity.
CC (NB, the amino acid sequence of isoform c' was produced by
CC adaptation of the wild-type ob receptor sequence provided in Fig 1
CC of the specification).
SO Sequence 892 AA.

```

alignment_scores: Quality: 61.50 Length: 111
Ratio: 0.992 Gaps: 8
Percent Similarity: 55.856 Percent Identity: 31.532

alignment_block: US-09-030-606-223 x W34260 ..

Align seg 1/1 to: W34260 from: 1 to: 892

```

17 AAAAAAATCTCT.....TCATTGAGAAAAATTAATCTTA..... 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
667 LysAsnAspSerLeuCySerValAlaGArgTyValValLysHisArgTh 683
50 .....GGGACT.....GATATGCTAATTAATGCGTAATTTA 80
683 rAlaHISAsnGlyThTrpSerGlnAspValGlyAsn.GlnThHisnLeu 699
81 ATWRTTKTGSGGCATTTCTTCAT.....TGCTTGA 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
700 ThrPheLeuTrpAlaGlnSerAlaHisThValThValLeuAlaIleAs 716
116 CAAGATTAATAATGTCGTGCCAAATTTGATTTATTTGAGACACTTCT 165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
716 nserIleGlyAlaSerleuValAsnPheAsnLeuThrPheSerTrpPROM 733
166 TATCAAAAGTAATGCTGCCAAGAGACTTAAGAAATTAGTAGTCC 215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4733 etSerLysVal.....AsnAlaValAlaGlnSerleuSerAlaTyPro 746

```

```

216 MTCAGTTGTTGGAGTGTGCTATTCTAAAGATTTGATTCCTGGAATG 265
747 LeuSerSer...SerCyValIleleu.SerTrpThrLeuSerProAsnA 762
266 ACAATTATATTTTAACCTTGTTGGTGGGGA 294
762 sPTySerleuLeuTyTrLeuValIleGlu 771

```

seq_name: A_Geneseq_36:W37338

seq_documentation_block:

```

ID W37338 standard; Protein: 894 AA.
AC W37338;
DT 07-MAY-1998 (first entry)
DE Ob protein receptor.
KW Ob protein receptor; obese phenotype detection; rat.
OS Rattus sp.
FH Key Location/Qualifiers
FT Misc_difference 805
FT /note= "encoded by TGT"
PN MO9741217-A1.
PD 06-NOV-1997.
PF 24-APR-1997; J01470.
PR 30-APR-1996; JP-134422.
PR (SAKA) OTSUKA PHARM CO LTD.
PI Iida M, Kodaira T, Murakami T, Shima K;
DR N-PSDB: T97156.
DR N-PSDB: T97156.
PT Ob protein receptor gene and variant having cytosine at base 806 -
PT for diagnosis of mutation(s) leading to development of obesity in
PT animals
PS Disclosure: Page 57-62; 71pp; Japanese.
CC This sequence represents the obesity protein receptor of the
CC invention. The gene sequence may be used for screening of warm blooded
CC animals (such as rats) for spontaneous mutations resulting in the variant
CC ob protein receptor gene which leads to animals having an obese
CC phenotype. These are useful in the study of mechanisms of obesity and the
CC development of anti-obesity drugs.
SO Sequence 894 AA.

```

alignment_scores: Quality: 61.50 Length: 111
Ratio: 0.992 Gaps: 8
Percent Similarity: 55.856 Percent Identity: 31.532

alignment_block: US-09-030-606-223 x W37338 ..

Align seg 1/1 to: W37338 from: 1 to: 894

```

17 AAAAAAATCTCT.....TCATTGAGAAAAATTAATCTTA..... 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
667 LysAsnAspSerLeuCySerValAlaGArgTyValValLysHisArgTh 683
50 .....GGGACT.....GATATGCTAATTAATGCGTAATTTA 80
683 rAlaHISAsnGlyThTrpSerGlnAspValGlyAsn.GlnThHisnLeu 699
81 ATWRTTKTGSGGCATTTCTTCAT.....TGCTTGA 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
700 ThrPheLeuTrpAlaGlnSerAlaHisThValThValLeuAlaIleAs 716
116 CAAGATTAATAATGTCGTGCCAAATTTGATTTATTTGAGACACTTCT 165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
716 nserIleGlyAlaSerleuValAsnPheAsnLeuThrPheSerTrpPROM 733
166 TATCAAAAGTAATGCTGCCAAGAGACTTAAGAAATTAGTAGTCC 215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
733 etSerLysVal.....AsnAlaValAlaGlnSerleuSerAlaTyPro 746
216 MTCAGTTGTTGGAGTGTGCTATTCTAAAGATTTGATTCCTGGAATG 265

```



```

      ::::::::::::::::::::
747 LeuSerSer...sercysvalileu.setrprhrluSerProAsnA 762
266 ACAATATATTTAACTTGTGGGGAA 294
      ::::::::::::::::::::
762 sPTySerleuLeuTyrlleuValileu 771

```

seq_name: A_Geneseq_36:W37337

seq_documentation_block:

```

ID W37337 standard; Protein; 894 AA.
AC W37337.
DE 07-MAY-1998 (first entry)
DE Ob protein receptor.
KW Ob protein receptor; obese phenotype detection; rat.
OS Rattus sp.
FH Key Location/Qualifiers
FT Misc_difference 805 /note="encoded by TGT"
FT MO9741217-A1.
PD 06-NOV-1997.
PE 24-APR-1997; J01470.
PR 30-APR-1996; JP-134422.
PA (SARA ) OTSUKA PHARM CO LTD.
PI Iida M, Kodaira T, Murakami T, Shima K;
DR WPI; 97-549722/50.
DR N-PSDB; T97149.
PT Ob protein receptor gene and variant having cytosine at base 806 -
PT for diagnosis of mutation(s) leading to development of obesity in
PT animals.
PS Claim 1; Page 49-54; 71pp; Japanese.
CC This sequence represents the obesity protein receptor of the
CC invention. The gene sequence may be used for screening of warm blooded
CC animals (such as rats) for spontaneous mutations resulting in the variant
CC Ob protein receptor gene which leads to animals having an obese
CC phenotype. These are useful in the study of mechanisms of obesity and the
CC development of anti-obesity drugs.
SQ Sequence 894 AA;

```

alignment_scores:

```

Quality: 61.50 Length: 111
Ratio: 0.992 Gaps: 8
Percent Similarity: 55.856 Percent Identity: 31.532

```

alignment_block:

US-09-030-606-223 x W37337 ..

Align seg 1/1 to: W37337 from: 1 to: 894

```

17 AAAAAAATCTCT.....TCATCAGAAAAATATCTTA..... 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
667 LysAsnAspSerLeuCySerValAlaArgTyValValLysHisArgTh 663
50 .....GGGACT.....GATATGGTAATTAATGTCATTTA 80
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
683 rAlaHisAsnGlyThrTrpSerGlnAspValGlyAsn.GlnThrAsnLeu 699
81 ATWRRRTTITGGGGCATTTCCCTTACAT.....TGCTTGA 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
700 ThrPheLeuTrpAlaGlnSerAlaHisThrValThrValLeuAlaIleAs 716
116 CAAGATTAAATGTCTGCCAAAGAGTCTAAGAAATTAAGTAGTCTCC 215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
716 nSerIleGlyAlaSerLeuValAsnPheAsnLeuThrPheSerTrpPro 733
733 etSerIlyVal.....AsnAlaValGlnSerLeuSerAlaTyrrPro 746
216 MTACATTGTTGGAGTGTCTATCTTAAGATTGATTGATTCCTGAAG 265
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
747 LeuSerSer...sercysvalileu.setrprhrluSerProAsnA 762

```

```

266 ACAATATATTTAACTTGTGGGGAA 294
      ::::::::::::::::::::
762 sPTySerleuLeuTyrlleuValileu 771

```

seq_name: A_Geneseq_36:W34258

seq_documentation_block:

```

ID W34258 standard; Protein; 895 AA.
AC W34258.
DE 11-MAY-1998 (first entry)
DE Rat Ob receptor isoform f.
KW Ob receptor isoform f; obesity; leptin; ligand; rat.
OS Rattus sp.
FH Key Location/Qualifiers
FT Misc_difference 805 /note="encoded by TGT"
FT MO9742340-A1.
PD 13-NOV-1997.
PE 02-MAY-1997; U07521.
PR 24-MAY-1996; GB-010995.
PR 06-MAY-1996; US-016899.
PA (MERI ) MERCK & CO INC.
PI Caskey CT, Hess JW, Hey P, Phillips MS;
DR WPI; 97-558993/51.
DR N-PSDB; T93104.
PT New isoform(s) of the Ob (leptin) receptor - used for identifying
PT specific binding ligands, potentially useful for study, prevention
PT and treatment of obesity.
PS Disclosure: Page -; 34pp; English.
CC This protein comprises isoform f of the rat Ob receptor. The Ob
CC receptor has numerous isoforms resulting from alternative splicing;
CC 3 novel isoforms, designated f, g and c' are disclosed (see
CC W34258-60). Isoform f differs from the 1162-residue wild-type
CC protein (see W34257) in that after Lys-889 there are only 6 amino
CC acids. In the cDNA (see T93104), these codons are then followed by
CC a stop codon. The f isoform can be expressed in host cells,
CC particularly E. coli, yeast or mammalian cells. It is used to
CC identify specific binding ligands. Agonists, antagonists and
CC ligand mimetics can be identified that are potentially useful in
CC the study, prevention and treatment of obesity.
CC (NB. the amino acid sequence of isoform f was produced by
CC adaptation of the wild-type Ob receptor sequence provided in Fig 1
CC of the specification).
SQ Sequence 895 AA;

```

alignment_scores:

```

Quality: 61.50 Length: 111
Ratio: 0.992 Gaps: 8
Percent Similarity: 55.856 Percent Identity: 31.532

```

alignment_block:

US-09-030-606-223 x W34258 ..

Align seg 1/1 to: W34258 from: 1 to: 895

```

17 AAAAAAATCTCT.....TCATCAGAAAAATATCTTA..... 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
667 LysAsnAspSerLeuCySerValAlaArgTyValValLysHisArgTh 683
50 .....GGGACT.....GATATGGTAATTAATGTCATTTA 80
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
683 rAlaHisAsnGlyThrTrpSerGlnAspValGlyAsn.GlnThrAsnLeu 699
81 ATWRRRTTITGGGGCATTTCCCTTACAT.....TGCTTGA 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
700 ThrPheLeuTrpAlaGlnSerAlaHisThrValThrValLeuAlaIleAs 716
116 CAAGATTAAATGTCTGCCAAAGAGTCTAAGAAATTAAGTAGTCTCC 215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
716 nSerIleGlyAlaSerLeuValAsnPheAsnLeuThrPheSerTrpPro 733
733 etSerIlyVal.....AsnAlaValGlnSerLeuSerAlaTyrrPro 746
166 MTACAAAAGTAATGCTGCCAAAGAGTCTAAGAAATTAAGTAGTCTCC 215
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
747 LeuSerSer...sercysvalileu.setrprhrluSerProAsnA 762

```

```

216  MCAATTGTTGGATCGATCTATAAAGATTTGATTCCTGGATG 265
      :::::||||:|||||:|||||:|||||:|||||:|||||
747  LeuSerSer...SerCysValIleLeu.SerTrpIleuSerProHisnA 762
      :::::||||:|||||:|||||:|||||:|||||:|||||
266  ACAATTATATTTTAACCTTTGGTGGGAGAA 294
      ||:::|||||:|||||:|||||:|||||
762  sPrpyrSerIleuLeuValIleGlu 771
      :::::||||:|||||:|||||:|||||:|||||
seq_name: A_Geneseq_36_W34259

```

seq. documentation block:
ID: W34259 standard; Protein: 1015 AA.
AC W34259;
DT 11-MAY-1998 (first entry)
DE Rat ob receptor isoform g.
KW Ob receptor isoform g; obesity; leptin; ligand; rat.
OS Rattus sp.
PN M09742340-A1.
PD 13-NOV-1997.
PF 02-MAY-1997; 007521.
PR 24-MAY-1996; GB-010995.
PR 06-MAY-1996; US-016899.
PA (MERI) MERCK & CO INC.
PI Caskey CT, Hess JW, Hey P, Phillips MS;
WP1: 97-358993/51.
PT New isoform(s) of the Ob (leptin) receptor - used for identifying
PT specific binding ligands, potentially useful for study, prevention
PT and treatment of obesity
PS Disclosure: Page -: 34pp: English.
CC This protein comprises isoform g of the rat ob receptor. The ob
CC receptor has numerous isoforms resulting from alternative splicing;
CC 3 novel isoforms, designated f, g and c' are disclosed (see
CC W34259-60). Isoform g is much shorter than the 1162-residue
CC wild-type sequence (see W34257). 18 N-terminal amino acids are
CC spliced to a large fragment of the wild-type molecule, beginning at
CC Pro-166. The isoform then extends for the remainder of wild-type
CC ob receptor molecule. Alternatively, it could contain another
CC isoform variation, such as isoform a, b, c, d, e or f. The g
CC isoform can be expressed in host cells, particularly E. coli, yeast
CC or mammalian cells. It is used to identify specific binding
CC ligands. Agonists, antagonists and ligand mimetics can be
CC identified that are potentially useful in the study, prevention and
CC treatment of obesity.
CC (NN. The amino acid sequence of isoform g was produced by
CC adaptation of the wild-type ob receptor sequence provided in Fig 1
CC of the specification).
SQ Sequence 1015 AA;

```
alignment_scores:      61.50      Length:      111
                      Quality:
                      Ratio:      0.992      Gaps:      8
Percent Similarity:    55.856      Percent Identity: 31.532
```

alignment_block:

Align seg 1/1 to: W34259 from: 1 to: 1015

17 AAAAAACAATCT TCATGACAAAAATATCTTA 49
520 LysAsnAspSerLeuScySerValaIaIaGlyTyrValValIaLysHisAlaGly 536
50GGGACTGATATGGTAATATATGTCACATTA 80
536 ValHisAsnGlyThrTyrSerGlnAspValaIaGlyAsn.GlnThrAsnLeu 555
81 ATATRTTGTGGGCAATTCCTTACAT TGTCTGGA 115
553 ThrPheLeuTyrIaIaGlnSerAlaHisHisThrValThrValLeuAlaIaLeu 566
115 CAAGATTAATAATCTCTGTCGCCAAAAATTTTGTATTATTATTTGGAGACTTCT 165

[illegible]

seq_name: A_Geneseq_36:W23399

```
seq_documentation_block:
```

AC W23399;

DE Rat ob receptor (fa mutant

KW fatty; fa mutation; therapy.

FH	Key	Location/Qualifiers
75	7-4-3	1 00

FT	Label = Sig_peptide
20	1100

FT	FE	Denat	label= Mat_protein
30	0.30		

```

FT      /label= Extracellular_domain
      000 000

```

```
FT      /label= Transmembrane_domain
FT      /label= CT
```

Domain	860. .1162
FT	
EM	

FT	Misc_difference	269
FT		

PN W09731015-A1
PD 38-217C-1007

PF 18-FEB-1997; U02397.
 DP 25-2 DP-1006; CP-008473

PR 22-FEB-1996; US-090405
 PR 22-MAR-1996; US-013060

PA (MERI) MERCK & CO INC
 PI Cackow CT Host IV

DR WPI; 97-435085/40.
DE N-PSNB; T64063

PT	Rat wild-type a
PM	1400+14100+100

PT obesity
PS C1a1w 4. Page --: 3500: E007146b

CC This protein comprises the rat
CC (OB-P) amino acid sequence

CC (see T64962) obtained by PCR amplifications. The sequence of CC at only 1 position from the wild-type (401 bp) OB-P sequence

CC the α OB-R mutant receptor has proline at amino acid 269 rather
CC than glutamine (see W33399). This is a result of an A to C

transversion in the *ta* OB-R gene (see T64962). Every tissue examined in an *fa/fa* Zucker rat was homozygous for this mutation

CC The wild-type and mutant OB-R can be expressed in prokaryotic or
CC eukaryotic host cells. There are no useful in assays to identify

putative ligands, which may be receptor agonists or antagonists or bind with little effect and which may be useful in understanding

CC obesity and for its prevention and treatment.

CC /NB the sequence for the fa OB-R was obtained by adapting the

CC sequence of wild-type OB-R provided in Fig 1 of the specification).

Sequence 1162 AA

alignment_scores:

Quality:	61.50	Length:	111
----------	-------	---------	-----

PT New isoform(s) of the Ob (leptin) receptor - used for identifying
 PT specific binding ligands, potentially useful for study, prevention
 PT and treatment of obesity
 PS Disclosure: Fig 1; 34pp; English.
 CC This protein sequence comprises wild-type (lean) rat ob receptor.
 CC The ob receptor has numerous isoforms resulting from alternative
 CC splicing; 3 novel isoforms, designated f, g and c' are disclosed
 CC (see W34258-60). These isoforms also apply to mouse and human ob
 CC receptors. The isoforms can be expressed in host cells,
 CC particularly E. coli, yeast or mammalian cells. They are used to
 CC identify specific binding ligands. Agonists, antagonists and
 CC ligand mimetics can be identified that are potentially useful in
 CC the study, prevention and treatment of obesity.
 SQ Sequence 1162 AA;

alignment_scores:
 Quality: 61.50 Length: 111
 Ratio: 0.992 Gaps: 8
 Percent Similarity: 55.856 Percent Identity: 31.532

alignment_block:
 US-09-030-606-223 x W34257 ..

Align seg 1/1 to: W34257 from: 1 to: 1162

```

17 AAAAACAATCT.....TCATCGAANAATATCTTA..... 49
|||||:||||| ||| |||:|||||:
667 LysAsnAspSerLeuCySerValAlaGArgTyValValLysHisArgTh 683
50 .....GGGACT.....GATATGGTAATATGTCATTTA 80
|||||:||||| ||| |||:|||||:
683 rAlaHisAsnGlyThrTirPserGlnAspValGlyAsn.GlnThrAsnLeu 699
81 ATWRTTITGGGGCATTCCTACAT.....TGCTTGA 115
|||||:||||| ||| |||:|||||:
700 ThrPheLeuTirPAlaGlnSerAlaHisThrValThValLeuAlaIleAs 716
116 CAAGATTAAATGCTGTGCGAANAATTTGATTTATTTGGAGACTTTC 165
|||||:||||| ||| |||:|||||:
716 nSerIleGlyAlaSerLeuValAsnPheAsnLeuThrPheSerTirProX 733
166 TATCAAAAGTAATGCTGCCAAGAGATCTAAGCAATAGTAGTCTCC 215
|||||:||||| ||| |||:|||||:
733 eSerIerYVal.....AsAlaValGlnSerLeuSerAlaIerPro 746
216 MTCACCTGTTGGAGTGTCTATCTTAAAGATTTGATTCTCTGGAATG 265
|||||:||||| ||| |||:|||||:
747 LeuSerSer...SerCyValIleLeu.SerTirPThrLeuSerProAsnA 762
266 ACATATATATTTAACTTGGTGGGGAA 294
|| |||:||||| |||
762 spTyIerLeuLeuTyIerLeuValIleGlu 771

```

seq_name: A_Geneseq_36:W75958

seq_documentation_block:

ID W75958 standard; Protein; 200 AA.
 AC W75958.
 DT 11-DEC-1998 (first entry)
 DE Mouse cell surface protein.
 KW Mouse; cell surface protein; thymocyte; lymphocyte; cell adhesion;
 KW signal transduction; autoimmune disorder; allergy; diagnosis;
 KW mitogen-stimulated.
 OS Mus sp.
 PN W09838216-A1.
 PD 03-SEP-1998.
 PF 27-FEB-1998; J00837.
 PR 26-FEB-1998; JP-062217.
 PR 27-FEB-1997; JP-062290.
 PA (NIBS) JAPAN TOBACCO INC.
 PI Tamatani T, Tezuka K;
 DR WPI: 98-481144/41.

DR N-PSDB: V53200.
 PT Cell surface molecule expressed in thymocytes and lymphocytes and -
 PT mediating signal transmission and cell adhesion, and antibodies to
 PT it useful in treatment of auto-immune and allergic disorders.
 PS Claim 9; Page 110-112; 149pp; Japanese.
 CC The present sequence represents a mouse cell surface protein which is
 CC expressed by thymocytes and by mitogen-stimulated lymphocytes. The cell
 CC surface protein induces adhesion of mitogen-stimulated lymphocytes to
 CC antibodies recognising the cell surface protein. These antibodies also
 CC produce an increase in peripheral blood lymphocytes in the presence of
 CC an antibody recognising CD3 antigen. The cell surface protein contains
 CC the amino acid sequence FDDPPF in its extracellular region and the
 CC sequence YWVF in its intracellular region. The cell surface protein can
 CC be used in the prevention and treatment of autoimmune and allergic
 CC diseases, and in the diagnosis and investigation of such disorders.
 SQ Sequence 200 AA;

alignment_scores:
 Quality: 57.50 Length: 73
 Ratio: 1.513 Gaps: 2
 Percent Similarity: 52.055 Percent Identity: 32.877

alignment_block:
 US-09-030-606-223 x W75958 ..

Align seg 1/1 to: W75958 from: 1 to: 200

```

151 TATTTGGAGACTTCTTATCAAAAGTAATGCTGCCAAGAGACTTAAGA 200
|||||:||||| ||| |||:|||||:
128 TyrLeuHisIleTyrGlnSerGlnLeuCySerGlnLeuLysLeuTirPle 144
201 ATTAGTAGTGTCCGTCATGTTGGAGTGTCTATCTTAAAGATT 250
|||||:||||| ||| |||:|||||:
144 uProValGlyLeuProAlaPheValValLeuLeuLeuGlyCysIleL 161
251 TGATTTCCGTAATGACATTTATTTAACTTGTGGGGGAANACTT 300
|||||:||||| ||| |||:|||||:
161 euIleIerP.....PheSerLysLysLysTyrGlySerSerVal 174
301 ATAGACACACACTTCTGATCTGTGATCTGTAAATTAATCTTTATTCG 350
|||||:||||| ||| |||:|||||:
175 HisAspProAsnSerGlnTyrMet.....PheMetAl 185
351 ACTGTTTGGACCATTAAG 369
|| ||| ||| |||
185 aAlaValAsnThrAsnLys 191

```

seq_name: A_Geneseq_36:R58854

seq_documentation_block:

ID R58854 standard; Protein; 383 AA.
 AC R58854;
 DT 13-APR-1995 (first entry)
 DE Fowlpox virus (FPV) genomic EcoRI fragment on vector 443-88.8.
 KW Homology vector 443-88.8.
 OS Fowlpox virus.
 PN W09418014-A.
 PD 01-SEP-1994.
 PF 28-FEB-1994; U01826.
 PR 26-FEB-1993; US-024156.
 PA (JABG) NIPPON ZEON KK.
 PA (SYTR) SYNTRON CORP.
 PI Cochran MD;
 DR WPI: 94-294007/36.
 DR N-PSDB: Q68935.
 PT New recombinant fowl pox virus for use in vaccines - contains
 PT genes expressing antigens of Newcastle disease virus and opt.
 PT infectious bronchitis virus
 PS Example; Page 54-55; 85pp; English.
 CC The homology vector 443-88.8 contains a 3.5 kb FPV genomic
 CC EcoRI fragment which maps to the approx. 5.5 kb overlap of
 CC FPV genomic fragments Sall C and PstI F. The ORF spans a

CC HpaI insertion site. This ORF shows no AA sequence homology
 CC to any known pox virus genes.
 SQ Sequence 383 AA;

alignment_scores:
 Quality: 58.00 Length: 100
 Ratio: 1.415 Gaps: 7
 Percent Similarity: 41.000 Percent Identity: 28.000

alignment_block:
 US-09-030-606-223 x R58654 ..

Align seg 1/1 to: R58854 from: 1 to: 383

```

99 TCCTTACATGCTGTGACAAAGATTAA...ATGCTGTGCCAAATTTTG 145
|||||
43 SerLeuHIScysPheaspArgSerIysGlyLeuaspGlnProIysThrPh 59
|||||
146 TATTTTATTTGGAGACTTCTTATCAAAAGTAATGTCGCAAGAGAGTCT 195
|||||
59 elleuProGlyIysIysSerasn..... 67
196 AAGGAATTAGTAGTCTCCCTCCACTGTTGGAGTGTCTATT..... 239
|||||
68 .....AsnSerIleIys.....LeuGluValAlaIleaspThr 78
240 ...CTAAAGATTGTGATTCTCGAATGCAATATATTATTTACTTTGG 286
|||||
79 TyrLysLysaspSerasphe..... 85
287 TGGGGGAANAAGTTATAGACACAGCTCTTCACTTGTGATCTGTAAT 336
|||||
86 .....SerTyr.....SerHisProCysGlnI 93
337 TAACTTTTATGCACTGT.....TTTGACCAT 365
|||||
93 lepheGlnPheCysValSerGlyAsnPheserGlyLysArgPheaspHis 109

```

seq_name: A_Geneseq_36:R58600

seq_documentation_block:
 ID R58600 standard; Protein; 383 AA.
 AC R58600;
 DT 14-APR-1995 (first entry)
 DE Fowlpox virus protein encoded by homology vector 443-88-8.
 KW Homology vector 443-88-8; recombinant fowlpox virus.
 OS Fowlpox virus.
 PN MO9419015-A.
 PD 01-SEP-1994.
 PE 28-FEB-1994; U002252.
 PR 26-FEB-1993; US-024156.
 PA (STR) SYNPRO CORP.
 PI Cochran MD;
 DR WPI: 94-294008/36.
 DR N-PSDB: 070574.
 PT New recombinant fowl pox viruses - useful as vaccines against
 PT fowl pox virus, Newcastle Disease virus and infectious
 PS laryngotracheitis virus.
 PS Example: Page 63-64; 97pp; English.
 CC Homology vector 443-88-8 contains a 3.5 kb fowlpox virus genomic
 CC EcoRI fragment and is useful for the insertion of foreign DNA into
 CC FPV. This EcoRI fragment maps to the approx. 5.5 kb overlap of FPV
 CC genomic fragments SalI C and PstI F. There is a unique HpaI site
 CC which is designated the 680 site. This site interrupts the ORF
 CC in Q70574 at AA 226. This ORF shows no AA sequence homology to any
 CC known pox virus genes.
 SQ Sequence 383 AA;

alignment_scores:
 Quality: 58.00 Length: 100
 Ratio: 1.415 Gaps: 7

Percent Similarity: 41.000 Percent Identity: 28.000

alignment_block:
 US-09-030-606-223 x R58600 ..

Align seg 1/1 to: R58600 from: 1 to: 383

```

99 TCCTTACATGCTGTGACAAAGATTAA...ATGCTGTGCCAAATTTTG 145
|||||
43 SerLeuHIScysPheaspArgSerIysGlyLeuaspGlnProIysThrPh 59
|||||
146 TATTTTATTTGGAGACTTCTTATCAAAAGTAATGTCGCAAGAGAGTCT 195
|||||
59 elleuProGlyIysIysSerasn..... 67
196 AAGGAATTAGTAGTCTCCCTCCACTGTTGGAGTGTCTATT..... 239
|||||
68 .....AsnSerIleIys.....LeuGluValAlaIleaspThr 78
240 ...CTAAAGATTGTGATTCTCGAATGCAATATATTATTTACTTTGG 286
|||||
79 TyrLysLysaspSerasphe..... 85
287 TGGGGGAANAAGTTATAGACACAGCTCTTCACTTGTGATCTGTAAT 336
|||||
86 .....SerTyr.....SerHisProCysGlnI 93
337 TAACTTTTATGCACTGT.....TTTGACCAT 365
|||||
93 lepheGlnPheCysValSerGlyAsnPheserGlyLysArgPheaspHis 109

```

seq_name: A_Geneseq_36:W10692

seq_documentation_block:
 ID W10692 standard; Protein; 383 AA.
 AC W10692;
 DT 05-MAY-1997 (first entry)
 DE Homology vector 443-88-8-encoded protein.
 KW Fowlpox virus; FPV; recombinant virus; vector; vaccine;
 KW Immunisation; Newcastle disease virus; NDV; poultry.
 OS Fowlpox virus.
 PN WO9640880-A1.
 PD 19-DEC-1996.
 PE 04-JUN-1996; U11187.
 PR 07-JUN-1995; US-484790.
 PA (STR) SYNPRO CORP.
 PI Cochran MD, Junker DE, Singer PA;
 DR WPI: 97-087060/08.
 DR N-PSDB: T48511.
 PT New recombinant fowlpox virus - contg. a foreign DNA sequence
 PT inserted into the fowlpox virus genome, used for the prodn. of
 PT vaccines.
 PS Example 1A; Page 94-95; 134pp; English.
 CC A polypeptide (W10692) that shows no homology to any known pox
 CC virus protein is encoded by an open reading frame found in a 1.5 bp
 CC fragment (T48511) of vector 443-88-8. This vector contains a 3.5
 CC kb fowlpox virus (FPV) genomic EcoRI fragment that maps to the
 CC approx. 5.5 kb overlap of FPV genomic fragments SalI C and PstI F
 CC and contains a unique HpaI site that interrupts the open reading
 CC frame. Vector 443-88-8 is useful for the insertion of foreign DNA
 CC into FPV. Insertion of an SfiI fragment (T48510) contg. Newcastle
 CC disease virus (NDV) hemagglutinin (HN) and fusion (F) genes
 CC yielded vector 502-26-22 (see also T48502-04), which can be used to
 CC insert NDV HN and F genes into FPV. The recombinant virus is used
 CC to deliver the vaccine antigens to poultry.
 SQ Sequence 383 AA;

alignment_scores:
 Quality: 58.00 Length: 100
 Ratio: 1.415 Gaps: 7
 Percent Similarity: 41.000 Percent Identity: 28.000

alignment_block:
US-09-030-606-223 x W10692 ..

Align seg 1/1 to: W10692 from: 1 to: 383

```

99 TCCTTACATGTCCTTGACAGATTAA...ATGCTGTGCCAAATTTTG 145
|||||
43 SerLeuHisCysPheAspSerLeuGlyLeuAspGlnProLysThrPh 59
|||||
146 TATTTTATTTGAGACCTCTATCAAAAGTATCTGCCAAAGAGACT 195
|||||
59 eileLeuProGlyLysTyrSerAsn..... 67
196 AAGCAATAGTACTGTTCCCTCCTACTGTTTGAGATGCTATT..... 239
|||||
68 .....AsnSerLeuLys.....LeuGluValAlaIleAspThr 78
|||||
240 ...CTAAAGATTTTGATTTTCCCTGGAATGACATTTATTTTAACTTTGG 286
|||||
79 TyrLysLysAspSerAspPhe..... 85
287 TGGGGGAAAGATATAGACACAGCTTCTGACTTGATCTGTAAT 336
|||||
86 .....SerTyr.....SerHisProCysGlnI 93
337 TAATCTTTATTCGACTTGT.....TTGACCAT 365
|||||
93 LepheGlnPheCysValSerGlyAsnPheSerGlyLysArgPheAspHis 109
|||||

```

seq_name: A_Geneseq_36:W01213

seq_documentation_block:

ID W01213 standard; Peptide: 119 AA.

AC W01213:

DE 16-MAR-1997 (first entry)

KW Flea; midgut; serine protease; PfSP19-119;

KW domestic animal; infestation; insecticide; protease-inhibitor;

OS Siphonaptera sp. location/Qualifiers

FT Key 69..71

PN W09611706-A1. /note="Conserved GWC sequence"

PD 25-APR-1996.

PR 18-OCT-1995; U14442.

PR 18-OCT-1994; US-326773.

PR 07-JUN-1995; US-485455.

PR 07-JUN-1995; US-482130.

PR 07-JUN-1995; US-485443.

PR 07-JUN-1995; US-484211.

PA (PARA-) PARAVAX INC.

PI (HEK-) HESKA CORP.

PI Aristen A, Dale B, Frank GR, Griewe RB, Heath A;

PI Hunter SW, Rushlow KE, Stiegler GL, Yamanaka M;

DR N-PSDB; T40858.

PT DNA encoding flea serine protease and aminopeptidase - useful in

PS Claim 71; Page 194; 24pp; English.

CC This sequence (PfSP19-359) represents a flea midgut serine protease

CC encoded by the PfSP19-359 gene, isolated from a flea cDNA library by

CC homology with conserved serine protease sequences (e.g. W01226).

CC The sequence contains a conserved GWC sequence, and shows homology

CC to cattle duodenase-1. The protein may be used in a vaccine for

CC protection of domestic animals from flea infestation, or in isolation

CC of protease-inhibitors, which may be used in controlled release

CC formulations to reduce the flea burden on and around the animal.

CC The inhibitors may be included in insecticidal compositions to

CC increase efficacy of other active agents, by reducing proteolytic

CC activity in the flea midgut.

CC Sequence 119 AA.

alignment_scores:

Quality: 56.50 Length: 43

Ratio: 2.093 Gaps: 1

Percent Similarity: 62.791 Percent Identity: 32.558

alignment_block:
US-09-030-606-223/rev x W01213 ..

Align seg 1/1 to: W01213 from: 1 to: 119

```

232 CACTCCAAACAGTGAAGGAGACACTAATTCCTTACACTTCTTGG 183
|||||
55 HisGlnThrGlnIleLysGlyGlyGluIleCysValAlaThrGlyTrpAl 71
|||||
182 CAGCATTACTTTGATBAGAGCTCCCAATTAATAAATTAATTTGGCA 133
|||||
71 YArgLeuGlyAlaAspGlnProValProAsnLysLeuGlnGlnLeuGln 88
132 CAGACATTTTATCTGTGTCAGCAATGT 104
|||
88 hr...IleAlaIleSerAspGlnLysCys 96

```

seq_name: A_Geneseq_36:P92108

seq_documentation_block:

ID P92108 standard; Protein: 637 AA.

AC P92108:

DE 14-FEB-1990 (first entry)

KW Human growth hormone receptor.

OS Homo sapiens.

PN US4857637-A.

PD 15-AUG-1989.

PF 12-JUN-1987; 061942.

PR 22-MAY-1985; US-737302.

PR 07-MAY-1986; US-861236.

PR 12-JUN-1987; US-061942.

PA (GENH) Genentech Inc.

PI Hammonds RG, Leung DW, Martin DW, Spencer SA, Wood WT;

DR WPI; 89-300419/41.

DR N-PSDB; N91325.

PT Modulating growth hormone receptor extracellular domain deriv. to raise

PT antisera.

PS Disclosure; Fig. 2a-c; 18pp; English.

CC An animal can be immunised against its growth hormone receptor by

CC vaccinating against a growth hormone receptor extracellular domain deriv.

CC predetermined to raise polyclonal antisera which affect the receptor as a

CC growth hormone agonist. This method enables continuous growth of target

CC tissues without frequent hormone admin.

CC Sequence 637 AA.

alignment_scores:

Quality: 57.50 Length: 53

Ratio: 1.983 Gaps: 2

Percent Similarity: 54.717 Percent Identity: 30.189

alignment_block:
US-09-030-606-223 x P92108 ..

Align seg 1/1 to: P92108 from: 1 to: 637

```

11 ACAAATAAAACAACTTCTTCATTCAGAAAAATTAATCTTAGGAGCTGATAT 60
|||||
232 SerLysGlnArgAsnSer..... 237
61 TGCTATATATGATTAATATATATATATATATATATATATATATATAT 110
|||||
238 .GlyAsnTyrGlyGlnPheSerGlnValLeu.....TyrValThrLeuP 252
111 CTGACAAAGATTAATATGCTGTGCCAAATTTGATTTATTTGAGAGA 160

```

252 roGlnMetSerGlnPheThrCysGluGluAspPheTyrPheProTirPleu 268
161 CTCTTATC 169
269 LeuIleIle 271

This Page Blank (uspto)

337 TAATCTTTATGCACTGT.....TTTGACAT 365
 : : : : :
 93 lepheGlnPheCysValSerGlyAsnPheserGlyLysArgPheAspHis 109

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US94-02252A-4

seq_documentation_block:
 ; Sequence 4, Application PC/TUS9402252A
 ; GENERAL INFORMATION:
 ; APPLICANT: Syntro Corporation, et al.
 ; TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses Thereof
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: John P. White
 ; STREET: 30 Rockefeller Plaza
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10112
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US94/02252A
 ; FILING DATE: 28-FEB-1994
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White Esq, John P
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212)977-9550
 ; TELEFAX: (212)664-0525
 ; TELEX: 422523
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 383 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US94-02252A-4

alignment_scores:
 Quality: 58.00 Length: 100
 Ratio: 1.415 Gaps: 7
 Percent Similarity: 41.000 Percent Identity: 28.000

alignment_block:
 US-09-030-606-223 x PCT-US94-02252A-4 ..

Align seg 1/1 to: PCT-US94-02252A-4 from: 1 to: 383

99 TCCTTACATTTGCTTGACAAGATTAA..ATGTCGTGTCGCAAAATTTTG 145
 : : : : :
 43 SerLeuHisCysPheAspArgSerLysGlyLeuAspGlnProLysThrPh 59
 : : : : :
 146 TATTTATTTGGAGACTTCTTATCAAAAGTAATGCTGCCAAGAAGACTCT 195
 : : : : :
 59 etLeLeProGlyLysTyrSerAsn..... 67
 : : : : :
 196 AAGGAATTAAGTAGTGTCCCTCCTGAGTGTGAGTGCTATT..... 239
 : : : : :
 68AnSerIleLys.....LeuGluValAlaIleAspThr 78
 : : : : :
 79 TyrLysLysAspSerAspPhe..... 85
 : : : : :
 240CTAAAGATTTTGATTTCTGGAATGACAATTAATTTTAACCTTGG 286
 : : : : :
 287 TGGGGGAAAAGATTATAGACACAGCTTCACCTTCGATACCTTGTAAT 336
 : : : : :
 86SerTyr.....SerHisProCysGlnI 93

337 TAATCTTTATGCACTGT.....TTTGACAT 365
 : : : : :
 93 lepheGlnPheCysValSerGlyAsnPheserGlyLysArgPheAspHis 109

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pep:PCT-US95-14442A-109

seq_documentation_block:
 ; Sequence 109, Application PC/TUS9514442A
 ; GENERAL INFORMATION:
 ; APPLICANT: Grieve, Robert B.
 ; APPLICANT: Rushlow, Keith E.
 ; APPLICANT: Hunter, Shirley Wu
 ; APPLICANT: Frank, Glenn R.
 ; APPLICANT: Heath, Andrew W.
 ; APPLICANT: Yamaka, Miles Yamanaka
 ; APPLICANT: Arfsten, Ann
 ; APPLICANT: Dale, Beverly
 ; TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
 ; TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
 ; TITLE OF INVENTION: INFESTATION AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
 ; MOLECULES, AND USES THEREOF
 ; NUMBER OF SEQUENCES: 119
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sheridan Rose & McIntosh
 ; STREET: 1700 Lincoln Street, Suite 3500
 ; CITY: Denver
 ; STATE: Colorado
 ; COUNTRY: USA
 ; ZIP: 80203
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/14442A
 ; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gary J. Connell
 ; REGISTRATION NUMBER: 32,020
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (303) 863-9700
 ; TELEFAX: (303) 863-0223
 ; INFORMATION FOR SEQ ID NO: 109:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 119 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US95-14442A-109

alignment_scores:
 Quality: 56.50 Length: 43
 Ratio: 2.093 Gaps: 1
 Percent Similarity: 62.791 Percent Identity: 32.558

alignment_block:
 US-09-030-606-223/rev x PCT-US95-14442A-109 ..

Align seg 1/1 to: PCT-US95-14442A-109 from: 1 to: 119

232 CACTCCAAAGAGTAGAGGAGACACTAATTCCTTAGACTTCCTTTGG 183
 : : : : :
 55 HisGlnThrGlnIleLysGlyGluIleCysLysAlaThrGlyIrrpGl 71
 : : : : :
 182 CAGCATTAATTTGATTAAGAGAGCTTCGAAATAAATAACAAATTTTGGCA 133
 : : : : :
 71 yArgLeuGlyAlaAspGlnProValProAsnLysLeuGlnGlnLeuGluT 88


```

TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC USES THEREOF
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/808,550
FILING DATE: 26-FEB-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1049-1-001 N
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 230 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: S. cerevisiae
STRAIN: Sce nfe-S
US-08-808-550-34

alignment_scores:
Quality: 51.50 Length: 43
Ratio: 1.776 Gaps: 2
Percent Similarity: 67.442 Percent Identity: 37.209

alignment_block:
US-09-030-606-223 x US-08-808-550-34 ..

Align seg 1/1 to: US-08-808-550-34 from: 1 to: 230

168 TAAAGATGATGCTGCGCAAGAGCTAGAGATGATGTTCCOMT 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5 SerIysIleuAlaProValaAspIleIleGlyGlySerSerIleProVa 21

218 CACT..TGTGGAGCTGCTCTATTCTTAAAGAATTTGATTCTCGAAT 264
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
21 IthValaIaserIyGcysGlyIleSerIyGlu.GlnIleSerProArg 37

265 GACAAATTAATTTTAACTTTGGTGGGG 291
|||||:|||||:|||||:|||||:|||||:|||||:
38 AspTyfArgLeuGlnValIleuLeuGly 46

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-219-237B-5

seq_documentation_block:
; Sequence 5, Application US/08219237B
; Patent No. 5874546
; GENERAL INFORMATION:
; APPLICANT: NAGATA, Shigekazu
; APPLICANT: ITOH, Naoto
; APPLICANT: YONEHARA, Shin
; TITLE OF INVENTION: DNA Coding for Human Cell Surface Antigen

```


PCT-US95-04018-49

Quantity:	21:00	4
Ratio:	1.759	2
Percent similarity:	65.909	
Percent identity:	31.818	

alignment_block:

US-09-030-606-223/rev x US-08-424-641B-2 ..

Align seg 1/1 to: US-08-424-641B-2 from: 1 to: 284

```

184 GGCAGCATTTTGTGATAGAGTCTCCAATAA.....ATACAAA 141
|||||  |||  ::  |||  |||  |||  |||  |||  |||  |||  |||
46 GLYLALALALEUPHEPHEGLUPROALPROGLINYSALAVALILEASVAS 62
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
140 TTTGGCAGACATTTTATCTTGTCAAGACATGTAAAGGAATGCCCC 91
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
62 pPhasnsersclueulleasncystyrargclmetlysaspnsprog 79
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
90 AMAAYAWATTAAATGACCATTAATTCACAT 59
::  |||  |||  |||  |||  |||  |||  |||  |||  |||
79 luginleullegluleu.....leuthrasn 87
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-162-081B-37

seq_documentation_block:

```

; Sequence 37, Application US/08162081B
; Patent No. 5824492
; GENERAL INFORMATION:
; APPLICANT: Hiles, Ian Donald; Fry, Michael John; Dhand, Ritu
; APPLICANT: Bala; Waterfield, Michael Derek; Parker, Peter
; APPLICANT: Joseph; Otsu, Masayuki; Panayotou, George; Volinla,
; APPLICANT: Stefano; Gout, Ivan Tarasovitch
; TITLE OF INVENTION: POLYPEPTIDES HAVING KINASE ACTIVITY,
; TITLE OF INVENTION: THEIR PREPARATION AND USE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felte & Lynch
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,081B
; FILING DATE: February 7, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00761
; FILING DATE: 13 April 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pasqualini, Patricia A.
; REGISTRATION NUMBER: 34,894
; REFERENCE/DOCKET NUMBER: LUD 5256
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-8200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1069 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-162-081B-37

```

alignment_scores:

```

Quality: 52.00 Length: 76
Ratio: 1.209 Gaps: 2
Percent Similarity: 56.579 Percent Identity: 25.000

```

alignment_block:

US-09-030-606-223 x US-08-162-081B-37 ..

Align seg 1/1 to: US-08-162-081B-37 from: 1 to: 1069

```

6 AACCAACAAACAAACAAATTCCTTCAGACAAATTAATCTTAGGAGCT 55
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
227 LysylsthrargsermetleuleuserSergluginleuylsleucysva 243
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
56 GATATTGGTAATATATGCTCAATTTAATWRTTRTKTGCGCATTTCCCTAC 105
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
243 lLeuglutyglnglyllystyrlleuleulysValcysglycysaspolut 260
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
106 ATGTCTTGACAGATTAATATGCTGTGCGCAAAATTT.....TGT 146
::  |||  |||  |||  |||  |||  |||  |||  |||  |||
260 yrphelenglulystyrProleuserclntyrlystyrlleargsercys 276
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
147 ATTTTATTGGAGACTTCTTATCAAAAGTAATGTCGCAAGAGAGCTA 196
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
277 lLeuleuleulargmetProashleuleuleuleulalalysgluser.. 292
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
197 AGCAATTAGTAGTGTTCCTCCMTCACTTGT 224
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
293 .LeutySerglInleuprometaspCys 301
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 12:39:06 ; Search time 2928.69 Seconds
(without alignments)
415.906 Million cell updates/sec

Title: US-09-030-606-223

Perfect score: 383
Sequence: 1 AAAACAACAACAACAACA.....ATTAGCTATATGTTAAAA 383

Scoring table: IDENTITY_NTC

Searched: 679419 seqs, 1590154680 residues

Database :

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_ph:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl1:*
10: gb_pl2:*
11: gb_pl3:*
12: gb_pl4:*
13: gb_pl5:*
14: gb_pl6:*
15: gb_pl7:*
16: gb_pl8:*
17: gb_pl9:*
18: gb_pl10:*
19: gb_pl11:*
20: gb_pl12:*
21: gb_pl13:*
22: gb_pl14:*
23: gb_pl15:*
24: gb_pl16:*
25: gb_pl17:*
26: gb_pl18:*
27: gb_pl19:*
28: gb_pl20:*
29: gb_pl21:*
30: gb_pl22:*
31: gb_pl23:*
32: gb_pl24:*
33: gb_pl25:*
34: gb_pl26:*
35: gb_pl27:*
36: gb_pl28:*
37: gb_pl29:*
38: gb_pl30:*
39: gb_pl31:*
40: gb_pl32:*
41: gb_pl33:*
42: gb_pl34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
c 1	378.4	98.8	197441	34	CNS00005	AL049779 Homo sapi

c	2	45.6	11.9	149764	10	HSAC002076	AC002076 Human BAC
c	3	45.2	11.8	207139	34	CEX1767	Z96049 Caenorhabdi
c	4	45.2	11.8	143092	36	CEX1767B	AL023828 Caenorhab
c	5	44.6	11.7	104308	8	AC006228	AC006228 Genomic s
c	6	44.6	11.6	82044	34	PFMAL1P2	AL031745 Plasmodi
c	7	44.6	11.6	151798	41	IP006960	U06960 Ichneumon p
c	8	44.2	11.5	151798	9	HS435C23	Z92844 Human DNA s
c	9	44	11.5	172007	42	AC006376	AC006376 Homo sapi
c	10	43.8	11.4	110680	8	AC006259	AC006259 Arabidops
c	11	43.8	11.4	36823	10	HSU209G1	Z68873 Human DNA s
c	12	43.8	11.4	21691	34	HS164F3	AL033422 Homo sapi
c	13	43.2	11.3	4284	36	DD020661	U02061 Dictyostell
c	14	42.8	11.2	96642	8	AC002291	AC002291 Arabidops
c	15	42.8	11.2	92612	8	ATAC003974	AC003974 Arabidops
c	16	42.8	11.2	274690	34	CEY60M9	AL022281 Caenorhab
c	17	42.8	11.2	167846	34	CEY66C5	Z98874 Caenorhabdi
c	18	42.8	11.2	24598	36	CEY24C2	Z68170 Caenorhabdi
c	19	42.8	11.2	12029	37	AE001400	AE001400 Plasmodi
c	20	42.6	11.1	300172	35	AC005308	AC005308 Plasmodi
c	21	42.4	11.1	143751	11	AC004160	AC004160 Homo sapi
c	22	42.4	11.1	142573	34	HSUJ753D5	AL049693 Homo sapi
c	23	42.4	11.1	929	36	DD1DCOF1	D37980 Dictyostell
c	24	42.4	11.1	3946	36	DD025143	U025143 Dictyostell
c	25	42	11.0	143299	10	HS81H13	AL023805 Human DNA
c	26	42	11.0	199606	35	AC004688	AC004688 Plasmodi
c	27	42	11.0	194410	35	AC005140	AC005140 Plasmodi
c	28	41.8	10.9	89281	8	ATAC006304	AC006304 Arabidops
c	29	41.8	10.9	97494	8	T15F16	AF076275 Arabidops
c	30	41.8	10.9	99902	10	HS59B16	AL032822 Human DNA
c	31	41.8	10.9	86829	36	PFMAL3P5	AL034556 Plasmodi
c	32	41.6	10.9	74342	7	AB020742	AB020742 Arabidops
c	33	41.4	10.8	107331	34	CEY49A10	Z93240 Caenorhabdi
c	34	41.4	10.8	312766	34	PFMALP3	AL035476 Plasmodi
c	35	41.4	10.8	37750	36	CEY504	Z50029 Caenorhabdi
c	36	41.2	10.8	12039	8	BTU25810	U25810 Bos taurus
c	37	41.2	10.8	79574	9	ATAC002340	AC002340 Arabidops
c	38	41.2	10.8	156909	9	AB020863	AB020863 Homo sapi
c	39	41.2	10.8	146813	11	AF121898	AF121898 Homo sapi
c	40	41	10.7	299081	34	AC006892	AC006892 Caenorhab
c	41	41	10.7	40168	36	CEYR06B10	AF040654 Caenorhab
c	42	40.8	10.7	83450	7	AB009054	AB009054 Arabidops
c	43	40.8	10.7	82594	8	ATAC005311	AC005311 Arabidops
c	44	40.8	10.7	154848	11	AC002992	AC002992 Homo sapi
c	45	40.8	10.7	160262	11	AC005242	AC005242 Homo sapi

ALIGNMENTS

RESULT 1
CNS00005/c
LOCUS
DEFINITION
Homo sapiens chromosome 14 clone bac R-1012A1 from RPCI-11 library,
WORKING DRAFT SEQUENCE, in ordered pieces.
ACCESSION
AL049779
NID
54760257
VERSION
AL049779.1 GI:4760257
KEYWORDS
HTG; HTGS; PHASE2.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 197441)
Genoscope.
AUTHORS
Direct Submission
TITLE
Submitted (05-MAY-1999) Genoscope - Centre national de sequence
JOURNAL
2, rue Gaston Cremieux - BP 191 91006 EVRY cedex - FRANCE E-mail :
sequence.tender@genoscope.cns.fr
COMMENT
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Location/Qualifiers

```

source
1..197441
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="bac R-1012A1 from RPOC-11 library"
BASE COUNT      57409 a 42215 c 42060 g 55757 t
ORIGIN

Query Match      98.8%; Score 378.4; DB 34; Length 197441;
Best Local Similarity 98.2%; Pred. No. 5.3e-61;
Matches 376; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY      1  AAACCAACCAACCAACCAACCAATCTTCATTCAGAAAAATATCTTAGGACGTGATAT 60
Db 50969  AAACCAACCAACCAACCAACCAACCAATCTTCATTCAGAAAAATATCTTAGGACGTGATAT 50910

OY      61  TGTGATATATGTCATTTAATWRTTRTITGGGGCATTCCTTACACTGTCTTGACAA 120
Db 50909  TGTGATATATGTCATTTAATAATATATTTGGGGCATTCCTTACACTGTCTTGACAA 50850

OY      121  TTAATATGCTGTGCCAAATATTTGATTTATTTTGAGACCTCTATCAAAAGTAAATGC 180
Db 50849  TTAATATGCTGTGCCAAATATTTGATTTATTTTGAGACCTCTATCAAAAGTAAATGC 50790

OY      181  TGCCAAAGAGACTCAAGGAATTAGTAGTTCCTCCMTCACTGTTGGAGTGTGATATTC 240
Db 50789  TGCCAAAGAGACTCAAGGAATTAGTAGTTCCTCCMTCACTGTTGGAGTGTGATATTC 50730

OY      241  TAAAGATTTGATTTCTCTGGAATGACAAATATATTTTAACTTTGGTGGGGGAAANAGTT 300
Db 50729  TAAAGATTTTATTTCTCTGGAATGACAAATATATTTTAACTTTGGTGGGGGAAAGAGTT 50670

OY      301  ATAGACCAACATCTTTCATCTTGATACTGTAAATTAATCTTTTATTCGACTTGTTTTG 360
Db 50669  ATAGACCAACATCTTTCATCTTGATACTGTAAATTAATCTTTTATTCGACTTGTTTTG 50610

OY      361  ACCATTAAAGCTATATGTTTAAAA 383
Db 50609  ACCATTAAAGCTATATGTTTAAAA 50587

```

RESULT	2
LOCUS	HSAC002076/c
DEFINITION	HSAC002076 149764 bp DNA PRI 12-MAY-1997
ACCSSION	Human BAC clone GS345D13 from 7q31-q32, complete sequence.
NID	AC002076 g2078461
VERSION	AC002076.1 GI:2078461
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Mammalia;
AUTHORS	Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE	1 (bases 1 to 149764)
JOURNAL	Maggi, L.
REFERENCE	The sequence of H. sapiens BAC clone GS345D13
TITLE	Unpublished (1997)
JOURNAL	2 (bases 1 to 149764)
REFERENCE	Waterston, R.
TITLE	Direct Submission
JOURNAL	Submitted (12-MAY-1997)
COMMENT	SUBMITTED BY: Genome Sequencing Center Department of Genetics Washington University St. Louis MO 63108, USA http://genome.wustl.edu/gsc mailto:saplense@wustl.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping phage sections once, or longer because we provide a small overlap

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION: The sequence of this clone was established as part of a mapping and sequencing collaboration between the NGR1 Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.ngri.nih.gov/DIR/GRB/CHR7> or send an E-mail to egreen@ngri.nih.gov

Mapping information for this clone was also provided by Dr. John D. McPherson, Department of Genetics/Genome Sequencing Center, Washington University School of Medicine.

SOURCE INFORMATION:
This clone is from the first BAC library from Genome Systems, Inc
(<http://www.genomesystems.com>).
Cell line: lymphoblastoid
Haplotypes: two
VECTOR: pBelOBAC

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is H.GS274A07, 200 bp overlap
actual start of this clone is at base position 1 of H.GS345D13
actual end is at 19827 of H.GS374A07. This cosmid lies in an
unanchored cluster of unknown orientation.

This clone contains STS SWSS4055 (NID:g1916534), SWSS2840 (NID:g113614), and SWSS1647 (NID:g113169).

FEATURES
source

```

/misc_feature      /organism="Homo sapiens"  
                    /db_xref="taxon:9606"  
                    /chromosome="7"  
                    /clone="GS345D13"  
                    /clone_1lb="GSBAC1"  
                    /map="7q31-q32"  
                    3138..4124  
repeat_region     /rpt_family="L1"  
                    4162..4237  
repeat_region     /rpt_family="L1"  
                    complement(4345..4638)  
misc_feature      /rpt_family="ALU"  
                    4769..5162  
                    /note="match to human EST AA039321 (NID:g15155599)  
                    2k39d04.s1"  
misc_feature      complement(5119..5551)  
                    /note="match to human EST AA039422 (NID:g1515699)  
                    2k39d04.r1"  
repeat_region     6029..6148  
                    /rpt_family="L1"  
misc_feature      6787..7179  
                    /note="match to human EST AA150894 (NID:g1722424)  
                    2147c03.s1"  
misc_feature      6788..7175  
                    /note="match to human EST AA135883 (NID:g1696857)  
                    212bd10.s1"  
misc_feature      6788..7201  
                    /note="match to human EST N33317 (NID:g1153716)  
                    yy08a02.s1"  
misc_feature      6802..7102  
                    /note="match to human EST H01232 (NID:g864165) yj31f03.s1"  
misc_feature      6909..7179  
                    /note="match to human EST H42319 (NID:g918371) y009g04.s1"  
misc_feature      complement(7074..7211)

```


* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES	Location/Qualifiers
source	1. .207139

BASE COUNT	a	c	g	t	
ORIGIN	68582	34583	35567	67606	801 others

Query Match	11.8%;	Score 45.2;	DB 34;	Length 207139;
Best Local Similarity	49.5%;	Pred. No. 1.8;		
Matches 140;	Conservative 1;	Mismatches 140;	Indels 2;	Gaps 1

97 TTTCCCTTACATTGTCCTTGACAAGATTAAATGTCCTGCGCCAAATTTGTATTTTATTG 13

Db 76011 TTGCTTATATGATTTAAAAAACATAAAAAATGCTTGAATAAAAGTTAGATTTTAAAG 76070

15/ GAGACTTCTATCAAAAGTATGCTGCCAAAGGAGCTTAAGGAATTAGTGTCCCM 216

DB / 80 / 1 GAGGCTTCCTGATACCGAATATCAATGCGAAAAAATTTCCCTGATTTTATA 76130

21 1C6C1G11G6AG1G1CIA1C1AMMAGH111GAT11C1GGA1GACCA11TA1AT 2

[illegible][illegible]

1000

[illegible]

JOURNAL Unpublished (1999)
 REFERENCE 2 (bases 1 to 104308)
 AUTHORS Ecker, J.R.
 TITLE Direct Submission
 JOURNAL Submitted (18-DEC-1998) Arabidopsis thaliana Genome Center,
 Department of Biology, University of Pennsylvania, 38th Street and
 Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
 REFERENCE 3 (bases 1 to 104308)
 AUTHORS Ecker, J.R.
 TITLE Direct Submission
 JOURNAL Submitted (11-APR-1999) Arabidopsis thaliana Genome Center,
 Department of Biology, University of Pennsylvania, 38th Street and
 Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
 On Apr 11, 1999 this sequence version replaced gi:4389503.
 COMMENT Location/Qualifiers
 FEATURES
 source 1..104308
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="1"
 BASE COUNT 33546 a 18958 c 18485 g 33319 t
 ORIGIN
 Query Match 11.7%; Score 45; DB 8; Length 104308;
 Best Local Similarity 48.5%; Pred. No. 2.4;
 Matches 114; Conservative 3; Mismatches 118; Indels 0; Gaps 0;
 QY 19 AATCATCTCTATTCAGAAAAATATCTTAGGAGCTGATATGTAATGTCAT 78
 DB 56700 AATTAATCTCAAGTTGAGCTTATTTACAGGCTGCACGACATTAATTAATT 56641
 QY 79 TAATRTTCTGAGGAGCATCTTCTTACATGCTTGACAGATTAATTAATGTCGCCAA 138
 DB 56640 TTGATTAATTAAGAGTAATCTTTAATTAATTAATTAATTAATTAATTAATTAAT 56581
 QY 139 AATTGTTGTTTATTTGAGAGCTTCTTATCAAAAGTAACTGCGCAAGAGAGTAAAG 198
 DB 56580 GACTCTCTCATTAATCAACGATGTTACTAATAAATCAATCAAAATATTTAACAT 56521
 QY 199 GAATAGTGTGCTGCCACACTGTTGGAGTGTCTTCTTCAAAAGATTTGA 253
 DB 56520 TATTTAAATATCAAAATTAATCAATTTTGTGATTTGCTTTTGCATGACTATA 56466
 RESULT 6
 PFMALIP2/C DNA HTG 09-APR-1999
 LOCUS Plasmidium falciparum chromosome 1 strain 307, WORKING DRAFT
 DEFINITION SEQUENCE, in unordered pieces.
 ACCESSION AL031745
 NID 94493855
 VERSION AL031745.3 GI:4493855
 KEYWORDS HTG; HTGS_PHASE1
 SOURCE malaria parasite P. falciparum.
 ORGANISM Plasmidium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmidium.
 1 (bases 1 to 82044)
 REFERENCE Bowman, S., Churcher, C., Harris, D., Lawson, D., Quail, M. and
 AUTHORS Barrell, B.
 TITLE Direct Submission
 JOURNAL Submitted (09-APR-1999) P. falciparum Genome Sequencing Consortium,
 The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
 CB10 1SA, UK
 COMMENT On Mar 24, 1999 this sequence version replaced gi:4455716.
 For more information about this sequence or the Malaria Project,
 see http://www.sanger.ac.uk/projects/P_falciparum. IMPORTANT: This
 sequence is unfinished and does not necessarily represent the
 correct sequence. Work on the sequence is in progress and the
 release of this data is based on the understanding that the
 sequence may change as work continues. The sequence may be
 contaminated with foreign sequence from E.coli, yeast, vector,
 phage etc.
 Order of segments is not known; 800 n's separate segments.

NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved
 FEATURES
 source 1..82044
 /organism="Plasmidium falciparum"
 /strain="3D7"
 /db_xref="taxon:5833"
 /chromosome="1"
 BASE COUNT 27301 a 7762 c 6747 g 29034 t 11200 others
 ORIGIN
 Query Match 11.6%; Score 44.6; DB 34; Length 82044;
 Best Local Similarity 46.7%; Pred. No. 3; Indels 0; Gaps 0;
 Matches 134; Conservative 2; Mismatches 151; Indels 0; Gaps 0;
 QY 1 AAACAAACAAACAAACAAACAAATTCATTCAGAAAAATTAATCTTAGGAGCTGATAT 60
 DB 46959 AAAAAAAAAAAAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATAT 46900
 QY 61 TCGTAATATGTCATTTAT 120
 DB 46899 AATTACAT 46840
 QY 121 TTAATATGTCGCGCAAAATTTGATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 180
 DB 46839 TTTTATGATTTTATTCAT 46780
 QY 181 TGCCAAAGAGAGCTAGAGATTAATAGTGTCCGTCAGTGTGGAGTGTGCTATTC 240
 DB 46779 TTAATTTGCTGAGAGAGATTAATTAAGAGTAAATTAACAAACAGATGATGAATAT 46720
 QY 241 TAAAGATTTTATTTCTGGAATGACATTAATTTATTTACCTTGT 287
 DB 46719 TATAAAGATGTTACACATGTTAATATATATATATATATATATATATATATATAT 46673
 RESULT 7
 IP006960
 LOCUS Ichneumon promissorius mitochondrial 16S rRNA gene, partial
 DEFINITION sequence.
 ACCESSION U06960
 NID 9463865
 VERSION U06960.1 GI:463865
 KEYWORDS
 SOURCE Ichneumon promissorius.
 ORGANISM Mitochondrion Ichneumon promissorius
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita;
 Ichneumonidae; Ichneumonidae; Ichneumoninae; Ichneumon.
 1 (bases 1 to 441)
 REFERENCE Dowton, M. and Austin, A.D.
 AUTHORS Molecular phylogeny of the insect order Hymenoptera: apocritan
 TITLE relationships
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (21), 9911-9915 (1994)
 MEDLINE 95024071
 REFERENCE 2 (bases 1 to 441)
 AUTHORS Dowton, M.
 TITLE Direct Submission
 JOURNAL Submitted (21-FEB-1994) Mark Dowton, Department of Crop Protection,
 Waite Campus, University of Adelaide, Adelaide, South Australia
 5064, Australia
 FEATURES
 source 1..441
 /organism="Ichneumon promissorius"
 /mitochondrion
 /db_xref="taxon:32393"
 /dev_stage="adult"
 <1..>441
 /note="large subunit ribosomal RNA"

OY	220	CTGTGGAGGCTCTATCTTAAGAATTTGATTCCTGGAAAGACAAATATATTTA	279
Db	66039	AATATTTTACTTTCTCTTGCGACATTTTCAGTTAGCCTTATGTAATCAATCAGT	66040
OY	280	ACTTGG	285
Db	66039	ACTTGG	66034
RESULT	10		
AC006259		DNA	PLN
LOCUS			28-DEC-1998
DEFINITION		Arabidopsis thaliana BAC F21J6 from chromosome V, containing KNAT3 and mapping near 60.5 cM, complete sequence.	
ACCESSION		AC006259	
NID		94063730	
VERSION		AC006259.1	GI:4063730
KEYWORDS		HTG.	
SOURCE		thale cress.	
ORGANISM		Arabidopsis thaliana	
		Eurariopsis, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.	
REFERENCE		1 (bases 1 to 110680)	
AUTHORS		Spiegel, L.A., de la Bastide, M., Habermann, K., Shah, R., Rodriguez, M., Shekher, M., Swaby, I., Matero, A., Preston, R., Vill, M.D., Schutz, K., Nascimento, L., Huang, E.N., Dedhia, N., Parnell, L.D. and McComble, W.R.	
TITLE		Arabidopsis thaliana BAC F21J6 from chromosome V, near 60.5 cM	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 110680)	
AUTHORS		Spiegel, L.A., de la Bastide, M., Habermann, K., Shah, R., Rodriguez, M., Shekher, M., Swaby, I., Matero, A., Preston, R., Vill, M.D., Schutz, K., Nascimento, L., Huang, E.N., Dedhia, N., Parnell, L.D. and McComble, W.R.	
TITLE		Direct Submission	
JOURNAL		Submitted (28-DEC-1998) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, NY 11724, USA	
COMMENT		BAC F21J6 is assigned to YAC C1C2E3 and maps to near 60.5 cM on the Lister & Dean RI map. Position 1 of F21J6, the T7 end, is oriented toward the telomere and position 110680, the Sp6 end, is oriented toward the centromere. For more information on the mapping, sequencing and annotation of F21J6, please see http://www.cshl.org/arabweb/F21J6-titlepage.html . A graphic view of our annotation is also available at this url. Gene models are built with exons predicted by GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), MZEF (http://www.cshl.org/genefinder) and GRAIL (http://compio.cornell.gov/tools/index.shtml) and with splice sites predicted by NetPlantGene (http://www.cbs.dtu.dk/netgene/obsmetgene.html). Alternate exons not used in building the gene models are presented on the web pages associated with F21J6. Genes are numbered according to the scheme BAC.gene.number. Typically, these numbers progress from 1 upwards as one moves from position 1 of the BAC. Protein sequences encoded by the genes are assigned to a functional category with the aid of similarity searches and comparison to the Prosite (http://expasy.hugue.ch/sprot/prosite.html) and Pfam (http://pfam.wustl.edu/) libraries. A description of these categories can be found at http://mnml.jac.mips.biochem.mpg.de/abi/ . Genomic repeats are typically located by REPEATX analysis and an attempt is made to classify the function of each repeat as either transposon, putative microsatellite, LINE, direct repeat, centromeric repeat, etc.	

If you have any questions or confirmatory or contradictory evidence concerning the annotation of F21J6, please direct email to Larry Parnell at parnell@cshl.org.

Location/Qualifiers

1. 110680

/organism="Arabidopsis thaliana"

gene
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="v"
/map="near 60.5 cM"
1. .717
/gene="F2J16.1"
/note="encodes putative P450; entire gene encoded on
T11H3, GenBank accession number AC005964; genomic copy of
EST Z33677; genomic copy of EST Z33963; gene model last
edited on 16 Dec 98"
/evidence-not_experimental

misc_feature
1. 24515
/note="overlap with T11H3, GenBank accession number
AC005964, from position 67404 to 91916"
/note="Region: overlap with T11H3"
join(<1. .11,103. .717)
/gene="F2J16.1"
/note="functional catalog ID=98"
/codon_start=3
/evidence-not_experimental
/product="putative P450"
/protein_id="AAC98443.1"
/db_xref="PID:g4063733"
/db_xref="GI:4063733"
/translation="ILENITLNGIDTSAQWTVMTYILISNPYMKAKQAEVREYIKN
KDDIIIEDIEELKLVKVKETFRVLPVLLIPREASKVYKIGYDIPKWTWHVNI
WAHNPNDPEALFIPERMNDIDYKGLNEFEELPFGSGRRMCPCIGMGALVHLT
LINLIRPDMKLPGMEVEDVDLEESYGLVCPKRPVQLIPVLTONS"
complement(1310. .1353)
/note="function-putative_microsatellite"
/rpt_type=" tandem
/rpt_unit="AT
4808. .6391
/gene="F2J16.2"
/note="encodes putative P450; gene model last edited on 16
Dec 98"
/evidence-not_experimental
join(4808. .5683,5777. .6391)
/gene="F2J16.2"
/note="functional catalog ID=98"
/codon_start=1
/evidence-not_experimental
/product="putative P450"
/protein_id="AAC98444.1"
/db_xref="PID:g4063734"
/db_xref="GI:4063734"
/translation="MSLITIVAFVFPSSMIYRIIRIKTKNLPPEPRRLPIITGNLHQ
LGSFPHKMSFLESTYPLGMSLKFGSVSTVASTPETERVKVLTDFDECCSRNMDP
ARVYVNLKDFSPYSKTYRWEVRKMTIVELVYTKARQSFGHTRKEVLAALDIKQAA
SLEPVRNKKMLKMLSGSVICRYVAFNLGSGKLENTYEEVIGTVEVAFPAADF
PVYRVIDIRITIGSKCEKLEFKAMDPEFQDSIKHLEDEIIKODIIDLLKMERGETT
LGEOLDRHQHKGITLIANLNGIDTSAQWTVMTYILISNPYMKAKQAEVREYIKN
DDIIEBDEIERQYIKMYIKETFRFRINPLIPREASKDVYKIGYDIPKWTWHVNI
AIIHNPNDPEALFIPERMDSQIDYKGLNEFEELPFGSGRRICPGIGMGALVHLT
LINLIRPDMKLPGEMKADVADVLEESYGLVCPKRIPIQLIPVLTONS"
23757. .24702
/gene="F2J16.6"
/note="mRNA sequence based on a comparison to cDNA U39646"
23757. .24702
/gene="F2J16.6"
/note="encodes ZFP3 zinc finger protein; genomic copy of
cDNA U39646"
/evidence-experimental
23757. .23810
/gene="F2J16.6"
/note="5'-UTR sequence based on a comparison to cDNA
U39646"
23811. .24518
/gene="F2J16.6"
/note="functional catalog ID=98"
/codon_start=1
/evidence-experimental
/product="ZFP3 zinc finger protein"

repeat_region

gene

CDS

5'UTR

CDS

OY	78	TTAATWRRTRKTGGGCATTCTTACATGTGTCGACAAGATTAAATCTCGTGCCA	137
Db	85303	TTTTGATAATTAGAAGAAAATCTTCATTAATTAATTTGACATTAAACATCTCCCA	85362
OY	138	AAATTTGATTAATTTATTTGGAGACTGCTTATCAAAAGTAATGCGCAGGAAGTCTAA	197
Db	85363	AATCTCTACCTTACCTACACGATTAATTAATTAATTAATTAATTAATTAATTAATTA	85422
OY	198	GGAATAGTAGTACTGTCCCMTCACCTGTGGAGTGTGCTATTAATTAAGATTGGATTTC	257
Db	85423	TTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	85482
OY	258	CTGGAATGACAAATTAATTAATTTAACTTTGGTGCGGGGAAANAGTTATA	303
Db	85483	CAATCCGATAGATTAATTTGATAGCATTTAACTTAACTAACAAATTAACA	85528

RESULT	11
LOCUS	HSU20961
DEFINITION	HSU20961 36823 bp DNA PRI 29-OCT-1997
ACCESSION	268873
NID	g1164913
VERSION	268873.1 GI:1164913
KEYWORDS	X.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS	Primates; Catarrhini; Homnidae; Homo.
JOURNAL	1: (bases 1 to 36823)
COMMENT	Ocell.C. Direct Submission Submitted (22-JAN-1996) E-mail enquiries: humquerry@sanger.ac.uk IMPORTANT: This sequence is the entire insert of clone cu20961. The true left end of clone U20961 is at 1 in this sequence. The true right end of clone Y1164A6 is at 7813. The true right end of clone U20961 is at 36823. This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre chromosome X mapping group. Further information can be found at http://www.sanger.ac.uk/HGP/chrX/ U20961 is from the Lawrence Livermore National Laboratory flow-sorted X chromosome cosmid library LLOXNC01. Location/Qualifiers 1..36823 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="X" /map="X" /clone="U20961" /clone_1lb="LLOXNC01" 768..853 /note="match: multiple ESTs" 1979..2268 /note="Alu repeat: matches 1..308 of consensus" 2269..2310 /note="21 copies of 2 mer 81 & conserved" 5628..7067 /note="12 copies of 120 mer 86 & conserved" 7294..7443 /note="2 copies of 75 mer 82 & conserved" 7822..7971 /note="2 copies of 75 mer 86 & conserved" 8041..8190 /note="2 copies of 75 mer 83 & conserved" 8118..8333 /note="2 copies of 108 mer 81 & conserved" 8597..8821 /note="3 copies of 75 mer 81 & conserved" 10080..10502 /note="match: 5' EST R79567 clone 145988" complement(10976..11105)

```

repeat__region /note="match: 3' EST R79568 clone 145988"
12107..12148
/note="L1 element fragment"
repeat__region 12233..12367
/note="L1 element fragment"
repeat__region 12577..12870
/partial
/note="Alu repeat: matches 308..1 of consensus"
16866..16892
/note="9 copies of 3 mer 96 & conserved"
repeat__region 17081..17123
/note="MER41 element fragment"
17583..17871
/partial
/note="Alu repeat: matches 308..1 of consensus"
18120..19739
/note="L1 element fragment"
repeat__region 20360..20484
/note="MSTC element fragment"
21043..21180
/note="L1 element fragment"
repeat__region 21181..21474
/partial
/note="Alu repeat: matches 308..1 of consensus"
21528..21647
/note="L1 element fragment"
repeat__region 21660..21848
/note="MLTIB element fragment"
21662..21713
/note="MLTIA element fragment"
21849..21876
/note="7 copies of 4 mer 96 & conserved; 14 copies of 2
mer 96 & conserved"
21849..21878
/note="5 copies of 6 mer 93 & conserved"
21927..22036
/note="MLTIB element fragment"
21966..22053
/note="MLTIA element fragment"
23206..23253
/note="MER7 element fragment"
23344..23570
/note="MER7 element fragment"
23586..23959
/note="MLTIB element fragment"
23588..23692
/note="MLTIC element fragment"
23745..23839
/note="MLTIC element fragment"
24903..25023
/partial
/note="Alu repeat: matches 148..10 of consensus"
26975..27080
/partial
/note="Alu repeat: matches 306..193 of consensus"
27432..27712
/note="Alu repeat: matches 1..308 of consensus"
28071..28220
/note="5 copies of 30 mer 81 & conserved"
28072..28203
/note="22 copies of 6 mer 83 & conserved; 33 copies of 4
mer 83 & conserved"
28072..28201
/note="55 copies of 2 mer 83 & conserved"
28074..28213
/note="5 copies of 28 mer 81 & conserved"
28266..28335
/note="MLTIC2 element fragment"
28389..28628
/note="MLTIC2 element fragment"
28872..28921
/note="2 copies of 25 mer 96 & conserved"
29191..29274

```

```

repeat_region      /note="3 copies of 28 mer 82 & conserved"
                   30879..31173
                   /partial
repeat_region      /note="Alu repeat: matches 308..1 of consensus"
                   32683..32857
                   /partial
repeat_region      /note="Alu repeat: matches 308..133 of consensus"
                   32860..32906
                   /partial
repeat_region      /note="Alu repeat: matches 47..1 of consensus"
                   34532..35401
                   /note="L1 element fragment"
repeat_region      35582..35988
                   /note="L1 element fragment"
repeat_region      36039..36219
                   /note="L1 element fragment"
repeat_region      36267..36563
                   /partial
BASE COUNT      10072 a 7172 c 8536 g 1123 t
ORIGIN

```

Query Match	11.4%: Score 43.8; DB 10; Length 36823;
Best Local Similarity 51.0%: Pred. No. 5.1;	
Matches 102; Conservative 0; Mismatches 98; Indels 0; Gaps 0;	
QY 113 TGACAGATTAAATGTCGTGCCAAAATTTTGATTTATTTTGAGACCTTTATCAA 172	
DB 31264 TAAATCAGATTGAGAACTAAATCTTAAAAACACAGAAATTCACAGTTAAATGTTATATAGA 31323	
QY 173 AGTATGCTGCCAAGAGAGCTTAAGAAATTAGTAGTGTTCGCCMCATCTTGTGGAGATG 232	
DB 31324 AATATATATAAAAATAAAAAGTACAGAGAAATTGAACAATGACATGAGTAAGATATTTAAATTT 31383	
QY 233 TGCATTCCTAAAGATTTTGATTTCTCGATGACAAATATATTTTAACTTTGGTGGGG 292	
DB 31384 TAAATTTTAAATTAATGCTGTTGTTGTGAGATTAACAGATTATTTTAAATTAGAAAGTATG 31443	
QY 293 AAANAAGTTATAGACCACAG 312	
DB 31444 GCCCAGATGATGGGACATAG 31463	

					RESULT	12
					HS164F3	
					LOCUS	
					DEFINITION	
					HS164f3	216991 bp DNA HTG 11-JUN-1999
					Homo sapiens chromosome X clone 164F3,	WORKING DRAFT SEQUENCE, In
					unordered pieces.	
					AL035422	
					NID	g5051831
					VERSION	AL035422.11 GI:5051831
					KEYWORDS	HTG; HTGS_PHASE1.
					SOURCE	human.
					ORGANISM	Homo sapiens
					Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
					Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
					1 (bases 1 to 216991)	
					Wilson,S.	
					Direct Submission	
					Submitted (11-JUN-1999) Wellcome Trust Genome Campus, Hinxton,	
					Cambridgeshire, CB10 1SA, UK. E-mail enquiries:	
					humangenesanger.ac.uk Clone requests: clonerequest@sanger.ac.uk	
					On Jun 12, 1999 this sequence version replaced gi:5050961.	
					IMPORTANT: This sequence is unfinished and does not necessarily	
					represent the correct sequence. Work on the sequence is in progress	
					and release of this data is based on the understanding that the	
					sequence may change as work continues. The sequence may be	
					contaminated with foreign sequence from E.coli, yeast, vector,	
					phage etc. Order of segments is not known. 800 n's separate	
					segments. Unfinished: dj164f3 Contig_ID: 03649 acc=AL035422	
					* NOTE: This is a 'working draft' sequence.	
					# NID: 216991 bp.	

Query Match	Similarity	11.3%	Score 43.2	DB 36	Length 4284
Best Local	Similarity	45.8%	Pred. No. 11		
Matches 135	Conservative	3	Mismatches 157	Indels 0	Gaps 0
OY	65	AATTATGGTCATTAATTAATTTTATTTGGGGGCAATTCCTTACATTCGTCTGCACAGATTAA	124		
DB	3297	AATTGAGGAGATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	3356		
OY	125	AATGCTGTGCCAAAATTTTGTATTTTATTTGGAGACTTCTTATCAAAAAGTAATGCTGCC	184		
DB	3357	GGTGTTCCTTTTCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	3416		
OY	185	AAAGGAGCTGATAGGATAGTAGTTCCTCCATCCTGTTGGATGGTGCATTCCTAA	244		
DB	3417	TATGAAATATAAAAAAATCCAAAATATTTGAAAAATCTATTTTCAAAAATATGATGCTTAT	3476		
OY	245	AGATTTGATTTCTTCGATGACAAATTAATTTTAACTTTGGTGGGGGAAANAAGTATAG	304		
DB	3477	TGATTAATTTTCTTCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3536		
OY	305	GACCACTCTTCACTTCGTGATCTTGTAAATTAATCTTTATTTGCACTGTTT	359		
DB	3537	CCCGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT	3591		
RESULT 14					
AC002291/c					
LOCUS	AC002291	96642 bp	DNA	PLN	15-JUL-1998
DEFINITION	Arabidopsis thaliana chromosome I BAC F2K20 genomic sequence, complete sequence.				
ACCESSION	AC002291				
NID	92477521				
VERSION	AC002291.1	GI:2477521			
KEYWORDS	HIG.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
	Euxariyta, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euhaydophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.				
REFERENCE	1 (bases 1 to 96642)				
AUTHORS	Federapfel,N.A., Palm,C.J., Conway,A.B., Kurtz,D.B., Conway,A.R., Au,M., Araujo,R., Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y., Oj,O., Osborne,B.I., Shinn,P., Sun,H., Toriumi,M., Vyotskala,V., Vyotskala,V.S., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.				
JOURNAL	Unpublished (1997)				
REFERENCE	2 (bases 1 to 96642)				
AUTHORS	Federapfel,N.A., Davis,R.W., Conway,A.B., Palm,C.J., Conway,A.R., Kurtz,D.B., Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y., Shinn,P., Sun,H., Oj,O., Osborne,B., Shen,Y.K., Toriumi,M., Vyotskala,V., Yu,G., Theologis,A. and Ecker,J.				
TITLE	Direct Submission				
JOURNAL	Submitted (05-JUN-1997) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA				
REFERENCE	3 (bases 1 to 96642)				
AUTHORS	Federapfel,N.A., Palm,C.J., Conway,A.B., Kurtz,D.B., Conway,A.R., Au,M., Araujo,R., Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y., Oj,O., Osborne,B.I., Shinn,P., Sun,H., Toriumi,M., Vyotskala,V., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.				
TITLE	Direct Submission				
JOURNAL	Submitted (07-OCT-1997) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA				
REFERENCE	4 (bases 1 to 96642)				
AUTHORS	Federapfel,N.A., Palm,C.J., Conway,A.B., Kurtz,D.B., Conway,A.R., Au,M., Araujo,R., Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y., Oj,O., Osborne,B.I., Shinn,P., Sun,H., Toriumi,M., Vyotskala,V., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.				
TITLE	Direct Submission				
JOURNAL	Submitted (04-FEB-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA				

COMMENT On Oct 7, 1997 this sequence version replaced gi|2251218.

e-mail for correspondence: arabseqsequence.stanford.edu
genes with similarity to proteins in the databases are described as
'putative', '-like' or 'similar to'. Genes that have EST
similarity but no significant protein similarity are described as
'unknown proteins'. Genes that are annotated based only on gene
prediction software are described as 'hypothetical proteins'.
The software programs used to predict genes include: Gtall
(Informatics Group, Oak Ridge National Laboratory,
<http://compbio.ornl.gov/section/index.html>), GENSCAN (Chris Burge,
<http://genome.stanford.edu/~chris/GENSCANW.html>) and NetPlantcane
(S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, <http://www.cbs.dtu.dk/Netplantene.html>). Location/Organism

FEATURES source

1..96642

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="I"

/clone="F22K20"

1692..3328

/gene="F22K20.1"

join(1692..2051,2148..2235,2316..2448,2542..3328)

/gene="F22K20.1"

/note="Similar to "tub" protein gpUB2468[2072162]; location of EST gb|U43528 and gb|AA395363"

/codon_start=-1

/protein_id="AAC00626.1"

/db_xref="PID:g2829918"

/db_xref="GI:2829918"

/translation=MSFRSLIVRDVRSIGLSRRSFDFKLSSLNKEGSGRGSYODSH EEDLVITIOETPMANLEPELLRLRYIKLSEESVPARRRVACASVCSPRMWDCKEI VOSEPVSGLKITFPVLSELQPGPRDAMTCEFKBSKSNTLYTLCLSPALLVEKGFEFL SARIIRRTTEERYVISMHADTIKRSSWTYGTRISNFIAGFTIIDYPDNYSINAIAR VOVGRISRRETSKYVSRPKPSGVSKRNNOVESYEILNGTCRGPRMHCAMSIPASSIAE GGVRPQPDIIVPRSIIIDESFRSTISSSKRIITYDSNDSSNAKFSDILEPLDEDOY VLGEKERNSPELVAKNPMPRHMQDLQCWLNRGTAVASKNFOLIANOPOPOPQ POPOPOLPPOPSGQTIDGDKIILLGFKNYKDMMTFDYPLSAFOFAICLSSFDT KLAE"

11246..12370

/gene="F22K20.2"

11246..12370

/gene="F22K20.2"

/note="Unknown protein; location of EST gb|AA394987"

/codon_start=-1

/protein_id="AAC00632.1"

/db_xref="PID:g2829924"

/db_xref="GI:2829924"

/translation="MESSPVNCLLRPDSLHOFISLPDIRIMICRSYCFENOLLTSOC FIEISTSRPLNLALRPHHHHSRSGHGANTPIYLHYVDPEONMFRENLDLF PFSPOVAASSGLVIYMGBSIDLAESSKTLCVNQLPTRFVKYLPOLGSAWSRGTVL VDSVNNRMVTTELAAATLYSCSYVANOMLRKSNMLPKRPRLKMSSSVFFLCDVGSW RSOQRLESCKLTINDLIITHHWCLERHEMDIDIIRPLLKRGNSDKLMTGGELS TFIAPNKCSITILRIDLDELWEAGEARMPLLEMVRCGOSESRTKVFGDGDRVFYSARK MGLIAMDCWGMRMGEGVGYADAELCRGFVPAKLTMF"

complemente(14545..15792)

/gene="F22K20.3"

/note="Highly similar to atExcl, extensin gene gb|U43627"

complemente(14671..15792)

/gene="F22K20.3"

/note="Highly similar to extensin (atExcl) protein gl|1167961"

/codon_start=-1

/product="extensin"

/protein_id="AAC00630.1"

/db_xref="PID:g2829922"

/db_xref="GI:2829922"

/translation="MASFILAFSLAFVOTANTFYRSSPPPVKHKHSPPPYKKSPPP PVKHSPSPYKSPPPVKHSPSPYKSPSPYKSPSPYKSPSPYKSPSPYKSPSPYK KHSPPPYKSPSPYKHKHSPPPYKSPSPYKHSBPPYKSPSPYKSPSPYKSPSPYK SPBPYKSPSPYKSPSPYKSPSPYKSPSPYKSPSPYKSPSPYKSPSPYKSPSPYKSP SPBPYKSPSPYKSPSPYKSPSPYKSPSPYKSPSPYKSPSPYKSPSPYKSPSPYKSP

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

```

Query Match      11.2%; Score 42.8; DB 8; Length 96642;
Best Local Similarity 46.0%; Pred. No. 6;
Matches 131; Conservative 3; Mismatches 151; Indels 0; Gaps 0;

gene
CDS
    OVIDINDEDDEDDDEAFALARSSPTPRPVGSSGWDILSGVYRPOGSSSS
    GYKPPPPPHPTSSSSSSPPDDASPOKSSSVIRLPPPPSRHNRQOFPFHHS
    SGSDSSTEGQTRILSLTSSSPQKEKREDLLFQVLEFATRRSKANNRSL
    complement(45417..46193)
    /gene="F22K20.8"
    complement(45417..46193)
    /gene="F22K20.8"
    /note="Hypothetical protein"
    /codon_start=1

Query 19 AACAATTCCTTCATTCAGAAAATTAATCTTAGGACGATATTGCTAATT 78
Db 10849 AAAATAGACATGAGATTAAAGATTATTATACAAATGCGTCGACATTTAATTAATT 10790
QY 79 TAATRRPTTTTGGGGCATTCCTTAATGCTGTGACAAAGATTAATGCTGTGCCAA 138
Db 10789 TTTTATATTTAGAAAAGAAAATCTCTATTTAAATTTGATTTTAACAACTTCTCAA 10730
QY 139 AATTTGATTTATTTTGGAGACTTCCTATCAAAAGTATGCTGCCAAAGAGACTTAAG 198
Db 10729 ATCTTCTACTTACACACAAATTAATTAATAAATTAACCTCCAAAATTTAAT 10670
QY 199 GAATTAGTAGTGTCCCTCCATCCTGTTTGGAGTGTCTATTCTAAAGATTTGATTTCC 258
Db 10669 CATTTAATTACTACAAAATTAATTCGTTTGTGATTTGATTTGGCAGATTAATCATG 10610
QY 259 TGGAAATGACATTAATTTTAACTTTGGTGGGGGAAAAGTATA 303
Db 10609 AAGCGCATAGATATTGATAGCATTTAATTAATTCACAAAATTAACA 10565

RESULT 15
ATAC003974/C
LOCUS ATAC003974 92612 bp DNA PLN 19-AUG-1998
DEFINITION Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence,
ACCESSION AC003974
NID 62914688
VERSION AC003974.1 GI:2914688
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eustoids II; Brassicales; Brassicaceae;
Arabidopsids.
REFERENCE 1 (bases 1 to 92612)
AUTHORS Rounsley,S.D., Lin,X., Ketchum,K.A., Crosdy,M.L., Brandon,R.C.,
Stykes,S.M., Kaul,S., Mason,T.M., Kelavange,A.R., Adams,M.D.,
Somerville,C.R. and Venter,J.C.
Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence
Unpublished
2 (bases 1 to 92612)
Rounsley,S.D. and Lin,X.
Direct Submission
Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org
3 (bases 1 to 92612)
Rounsley,S.D.
Direct Submission
Submitted (27-FEB-1998) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Feb 27, 1998 this sequence version replaced gi:2828768.
Address all correspondence to:
Steve Rounsley
The Institute for Genomic Research
9712 Medical Center Dr,
Rockville, MD 20850,
USA
COMMENT

```


e-mail: rounsley@tigr.org
 BAC clone F24L7 is from Arabidopsis chromosome II and close to the molecular marker TEN5.
 The orientation of the sequence is from SP6 to T7 end of the BAC clone.
 Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (available by anonymous ftp from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of Washington), GenScan (Chris Burge, <http://genomic.stanford.edu/~chris/GENSCAN.html>), and NetPlantGene (<http://www.cds.dtu.dk/netgene/cbsnetgene.html>), searches of the complete sequence against a peptide database and the Arabidopsis EST database at TIGR (<http://www.tigr.org/tdb/at/at.html>).
 Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of genomic sequence that are not annotated as genes but have predicted exons by GRAIL are annotated as misc features.

FEATURES

```

source
    location/Qualifiers
    1..92612
    /organism="Arabidopsis thaliana"
    /cultivar="Columbia"
    /db_xref="taxon:3702"
    /chromosome="II"
    /map="TEN5"
    /clone="F24L7"
    1..12476
    /note="overlap with BAC clone T21L14 (AC003033:84325..96801)."
    complement(331..458)
    /note="exon predicted by xgrail, quality good_shadowexon"
    1075..1117
    /note="exon predicted by xgrail, quality excellent"
    join(<1184..1650,1741..1906,1992..2042,2132..2381,2566..3919)
    /gene="F24L7.1"
    <1184..3919
    /gene="F24L7.1"
    /note="T21L14.21"
    join(1184..1650,1741..1906,1992..2042,2132..2381,2566..3713)
    /gene="F24L7.1"
    /codon_start=1
    /product="putative cyclin g-associated kinase"
    /protein_id="AAC04479.1"
    /db_xref="PID:g2914689"
    /db_xref="GI:2914689"
    /translation="MAKRGKPKAPKAPKAGLEGRLEIGNAKOVANVIAAGGSSVYLAQVNAHAKOYALKMTCNDESELELYAKETISVLSKSGHPVNTVLAHILDMGRKKEALLADFCGKSLVDLENRGAGYFEERQALITFRVCNAVFAHMCOSPRIAHARDKAENILSSDQWMLCDGFSVSTNKRIFERAEMGIEDNIRKTYTPYRAPEMWDLERRKEMISEKVIDMALGCLDFRICYFNADGESKQILNGNRIKIPSPKYSVEITDLIKEMLOASPERPDITOVKSOPFLIYSSOLRCKMMSLTGSSKATRPSAPRSPRPPPSGSGSGPGLGAFMATQAHAKTVSSEDNKMMPTEDPNSTKSESEVDRISHPKPPSPYRGAERGIQRNKDELTITISQKVTTPAANNMTRVSKDAFNSFVADPTITFDNENKPKGEALAEIQRKLDELKQTSSEKATIAKEKLSAIGRSORQELADKQTLASASASSPSPSSONQSPGMHMSSTPSVLFVLVHLSDTHTKYFINVYLLILFELMKRDKMEGTMEILODRSNMSTGSDTNMOPFDEDAKPYMESAGKNNNTINOSVTRPSKPSAAGTQGEFEPWGFETESFRAAATSAATASAGTGRMSGSGNSTSQRRGNSMRENOKTAQAPAGNAGF"
    complement(3972..4026)
    /note="exon predicted by xgrail, quality excellent_shadowexon"
    join(<4325..4879,4961..4993,5123..5175,5266..5305,5403..5480,5590..5697,5791..5870,5967..>6060)
  
```

```

gene
    /gene="F24L7.2"
    <4325..>6060
    /gene="F24L7.2"
    /note="T21L14.22; predicted by genefinder"
    join(<4325..4879,4961..4993,5123..5175,5266..5305,5403..5480,5590..5697,5791..5870,5967..6060)
    /gene="F24L7.2"
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="AAC04480.1"
    /db_xref="PID:g2914690"
    /db_xref="GI:2914690"
    /translation="ASTPDPPNPMPNPNSVAVSTPYTASESPVTPQNTVTPSSQPOPDPASSYRAIAPLRHHPNONTITNPDIRNSNTYNSPHOPHDPSSLIYRPGSSGRFPFPRPNQNSNADVPGSSPGGYTRGVYGHQPTNSIDPNKQFRRAHPQNOSSPOLSGHKGVPHPFLQPLVSEVVEATPSPSILDSGKKARSDDALVLRKRVRITDEGASLYSLCRSWLRNGAHGIRIDMCLPLPVDKTESLPKDLVEALIEDEDESESVKHLSESDLRHIDRAKVARALREERLKRIRYARIALILPPGEQCRNE"
    complement(6589..6697)
    /note="exon predicted by xgrail, quality excellent_shadowexon"
    6796..6918
    /note="exon predicted by xgrail, quality marginal"
    complement(6828..7013)
    /note="exon predicted by xgrail, quality marginal_shadowexon"
    7708..7972
    /note="exon predicted by xgrail, quality excellent"
    complement(7752..7844)
    /note="exon predicted by xgrail, quality excellent_shadowexon"
    complement(7914..7992)
    /note="exon predicted by xgrail, quality excellent_shadowexon"
    complement(8108..8225)
    /rpt_family="TAAAn"
    complement(8448..8593)
    /note="exon predicted by xgrail, quality excellent_shadowexon"
    complement(9131..9157)
    /rpt_family="AT-rich"
    complement(9372..9408)
    /note="exon predicted by xgrail, quality marginal_shadowexon"
    join(<9504..9807,9985..10457,10556..>11407)
    /gene="F24L7.3"
    <9504..>11407
    /gene="F24L7.3"
    /note="T21L14.23"
    join(9504..9807,9985..10457,10556..11407)
    /gene="F24L7.3"
    /codon_start=1
    /product="putative phosphate transporter"
    /protein_id="AAC04481.1"
    /db_xref="PID:g2914691"
    /db_xref="GI:2914691"
    /translation="MAKRGKVELNALDAKTOYHPTAIVIAAGMFFTDVADLFSISLWKLGRIVYHSDSSKRKPGTLPPNVAANGVAFGTLAGOLFGLMDKIGKRVYGITMMLGRISYSGSGSPGSHANGVAAATCFEPFGIGGDPPLSATIMSEYARKTBAFLAIPAMQGGIILAGCYLIVSSTPHARAPAYENDPNCSTYPOADYVRIVLSNGAIPALITTYRMRKPERIARTALVARTKQASDSKYLQVLDLAEEAOSNSNSNPPTGTLFREFRRRHGHLLDTTITWFLDLIAVSSNLQDITTAIGWIPAAETMAAHEFTVYSKAQTLALAGTVGWFVAFIDILRFFIOLMGFTMTIFMFAALIPYDMRRENRNIGLILYSLTMFPANGPVATFVVALETFPARLRSTCGISAAAGKAGAVGAGFGLYAAOSSDSEKTDAGYPGIGVRSILMLACVNFGLGVFLLVDESKGSLSEIRREDEQSGDPTVEMTVANGRRVAV"
    11653..11702
    /note="exon predicted by xgrail, quality marginal"
    complement(12069..12123)
    /note="exon predicted by xgrail, quality marginal_shadowexon"
    complement(<12375..>13376)
  
```

```
gene      /gene="F24L7.4"
          complement(<12375..>.13376)
          /gene="F24L7.4"
          /note="predicted by genefinder and genscan"
          complement(12375..13376)
CDS       /gene="F24L7.4"
          /note="hypothetical protein"
          /codon_start=1
          /protein_id="AAC04482.1"
          /db_xref="PID:G2914692"
          /db_xref="GI:2914692"
          /translation="WMKIAIAERTDLPRLVSTAIRASNSQMSDVERCFDYLAYLK
          GILNLSVKNLSHSHKYLPLESLRDHENPKRTEAHVLTSMKTEFYSSGONSSTCNKP
          NPLKTKVYKACSELEKNEOLSHGFAYLAKKETEYFLGKMKNEDESRVHEPREMK
          QIGDSKSPALMRTEKKKOSAPIGSKNPSGAGETGVIKKILKPDCCSVSPRP
          PLNMKKHPVKAFFENPKCIYLKKNSPENLELPEAKKSADYANAKGFLAKEPASTC
          VDIALLMKESIISTAIETTRIMETKLERLTKHKDKICNALALALHMRQTIINQO"
          complement(13364..13410)
```

repeat_region

```
Query Match      11.2%  Score 42.8; DB 8; Length 92612;
Best Local Similarity 46.0%  Pred. No. 6.1;
Matches 131; Conservative 3; Mismatches 151; Indels 0; Gaps 0;
```

```
OY  19  AACAATTCCTCATTGAGAAAATTAATCTTAGGGAGCTGATTTGTAATTATGTCATTT 78
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  53138 AATTAATCTGAGTGTGAGATTATTACAGTCCGCTACCAGCATTTAATAATTGTT 53139
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY  79  TAATWRTTRTKTGGGCAATTCCTTACATGTCCTGACAAGATTAAATGCTGTGCCAA 138
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  53138 TTTGATTAATTAGAAAGAAATCTTCTAATTAATTAATTGACATTTAACAATCTCCCAA 53079
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY  139  AATTTTGTATTTTATTTGAGACTTCTTATCAAAAGTAATGCTGCCAAGAGTCTAG 198
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  53078 ATCTTCTACCTTACTACATTAATTAATAATAAACTCCAAATATTTAATAAT 53019
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY  199  GAATTAGTAGTGTCCCMTCACCTGTGTTGAGTGTGTAATCTTAAAGATTGATTCC 258
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  53018 CATTTAATTAACAATAATTAATCATTTTGTGATATGCTTTGACATGATTAATCATC 52959
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY  259  TGGATGACAAATTAATTTACTTGTGGGGGGAANAGTTATA 303
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  52958 AAGCCGATGATTAATTGATGACATTTAATTAATCAAAATTAACA 52914
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: September 28, 1999, 12:41:49
Job time: 6189 sec

RESULT 2
 V61268 2
 ID V61268 standard; cDNA; 383 BP.
 AC V61268:
 DT 06-JAN-1999 (first entry)
 DE cDNA sequence of prostate tumour clone P5095.
 KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
 OS Homo sapiens.
 PN M08837093-A2.
 PD 27-AUG-1998
 PF 25-FEB-1998; U03492.
 PR 09-FEB-1998; US-020956.
 PR 25-FEB-1997; US-806099.
 PR 01-AUG-1997; US-904804.
 PA (CORI-) CORIXA CORP.
 PI Dillon DC, Xu J;
 DR WPI: 98-609886/51.
 PT Polypeptides comprising immunogenic portions of prostate proteins -
 PR used in a vaccine for the treatment of prostate cancer
 PS Claim 3; Page 124; 130p; English.
 CC The present sequence is a new DNA which encodes an immunogenic portion
 CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
 CC can be used as a vaccine for the treatment of prostate cancer. The DNA
 CC was identified by analysis of a subtracted cDNA library obtained by
 CC subtracting a prostate tumour cDNA expression library with a normal
 CC tissue cDNA library.
 SO Sequence 383 BP; 123 A; 51 C; 63 G; 140 T;

Query Match 99.2%; Score 380; DB 1; Length 383;
 Best Local Similarity 100.0%; Pred. No. 7.3e-77;
 Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAAAAACAACAAAAAATTCCTTCAGAAAAATTAATCTGGAGCTGAT 60
 DB 1 AAAAAACAACAAAAAATTCCTTCAGAAAAATTAATCTGGAGCTGAT 60
 OY 61 TGTATATATGTCATATTAATRTTGTGGGCAATTCCTTACATGCTTGAACA 120
 DB 61 TGTATATATGTCATATTAATRTTGTGGGCAATTCCTTACATGCTTGAACA 120
 OY 121 TTTAAATGCTGTGCCAAAATTTTGTATTTTGGAGACTTCTTAACAAAAGTAATGC 180
 DB 121 TTTAAATGCTGTGCCAAAATTTTGTATTTTGGAGACTTCTTAACAAAAGTAATGC 180
 OY 181 TGCCAAAGAGCTCTAGGAATTAAGTGTCCCTCACTGTTGGAGTGGCATTC 240
 DB 181 TGCCAAAGAGCTCTAGGAATTAAGTGTCCCTCACTGTTGGAGTGGCATTC 240
 OY 241 TAAAGATTTTGTATTCCTGGAATGACATTTATTTTAACTTGTGGGGAANAAGTT 300
 DB 241 TAAAGATTTTGTATTCCTGGAATGACATTTATTTTAACTTGTGGGGAANAAGTT 300
 OY 301 ATAGACACACAGTCTTCACTTGTGATCTTGAATTAATCTTTATTTGACCTGTTTGG 360
 DB 301 ATAGACACACAGTCTTCACTTGTGATCTTGAATTAATCTTTATTTGACCTGTTTGG 360
 OY 361 ACCATTAGCTATGTTTAAAA 383
 DB 361 ACCATTAGCTATGTTTAAAA 383

RESULT 3
 X20248_08
 Continuation (9 of 10) of X20248 from base 800001 (Borella burgdorferi polynucleotide s
 WP Sequence split into 10 fragments LOCUS X20248 Accession X20248

WP	Fragment Name	Begin	End
WP	X20248_00	1	110000
WP	X20248_01	100001	210000
WP	X20248_02	200001	310000
WP	X20248_03	300001	410000
WP	X20248_04	400001	510000
WP	X20248_05	500001	610000

WP	X20248_06	600001	710000
WP	X20248_07	700001	810000
WP	X20248_08	800001	910000
WP	X20248_09	900001	910715

Query Match 10.4%; Score 40; DB 1; Length 110000;
 Best Local Similarity 46.5%; Pred. No. 1.1;
 Matches 155; Conservative 4; Mismatches 155; Indels 9; Gaps 1;

OY 38 AAAATATCTTAGGAGCTGATATGTAATTAATGCTCAATTAATWRTTGTGGGCAT 97
 DB 14020 AAGAGTTCTATTGAATAATTTGACAGATATATAGTGTTTAGATGAGTTTAAAT 14079
 OY 98 TTCTTACATGCTGTGACAGATTAATAATGCTGTGCCAAAATTTTATTTATTTGG 157
 DB 14080 AAATTTATTTTGGCAATAAATAGATGATCAAGCATTTATATGTTCAATTTATCTGA 14139
 OY 158 AGACTTCTTATCAAA-----AGTATGCTGCCAAGAGCTTAAGAAATTAGTAG 208
 DB 14140 TTTATATTTTAAACCCAGAGCTTAAATAATTTGCTTAATGATTAAGAAACATATAGA 14139
 OY 209 TGTCCCATCACTGTTTGGAGTGTGCTATCTTAAAGATTTTATTCCTGGAATGACA 268
 DB 14200 TGTGATGCTCTGTAATGAAAGTGTGTGTAAGATATGATGATTTGAATTTGA 14259
 OY 269 ATTATATTTTAACCTTGTGGGGGGAANAAGTATAGACACAGCTTCATCTTGATAC 328
 DB 14260 AATTAATAATTTAATAATTAACAGCGTTGAAATCAGTAAATAAATAATTAATGGAACGC 14319
 OY 329 TTGTAATTAATCTTTTATTTGACCTGTTTGA 361
 DB 14320 AGAATATTTGACTTCTATTTTGTGTTATTA 14352

RESULT 4
 V21209_03/c
 Continuation (4 of 17) of V21209 from base 300001 (Methanococcus jannaschii circular
 WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

WP	Fragment Name	Begin	End
WP	V21209_00	1	110000
WP	V21209_01	100001	210000
WP	V21209_02	200001	310000
WP	V21209_03	300001	410000
WP	V21209_04	400001	510000
WP	V21209_05	500001	610000
WP	V21209_06	600001	710000
WP	V21209_07	700001	810000
WP	V21209_08	800001	910000
WP	V21209_09	900001	1010000
WP	V21209_10	1000001	1110000
WP	V21209_11	1100001	1210000
WP	V21209_12	1200001	1310000
WP	V21209_13	1300001	1410000
WP	V21209_14	1400001	1510000
WP	V21209_15	1500001	1610000
WP	V21209_16	1600001	1664976

Query Match 10.4%; Score 39.8; DB 1; Length 110000;
 Best Local Similarity 47.9%; Pred. No. 1.2;
 Matches 101; Conservative 4; Mismatches 106; Indels 0; Gaps 0;

OY 1 AAAAAACAACAAAAAATTCCTTCAGAAAAATTAATCTTAGGAGCTGATAT 60
 DB 41592 AAGATCCAGAGCAAGAAGTATGATAGTAGAAGATACATTAATGAGGAGAGATTG 41533
 OY 61 TGTATATATGTCATATTAATWRTTGTGGGCAATTCCTTACATGCTTGAACA 120
 DB 41532 TAGAATATGAACAGATATGATGATGTCAGCAGATTTGACCAAGACCAAGCAATTC 41473
 OY 121 TTTAAATGCTGTGCCAAAATTTTATTTATTTTGAAGACTTCTTAACAAAAGTAATGC 180

	RESULT
6	
ID	x20250 standard; DNA; 11309 BP.
AC	x20250
DT	04-MAY-1999 (first entry)
DE	Borrelia burgdorferi polynucleotide sequence #3.
KW	Borrelia burgdorferi, spirochete; bacterium; pathogen; Lyme disease.
KW	epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis.
KW	Infection; diagnosis; Characterisation; detection; ds.
OS	Borrelia burgdorferi.
PN	W09858943-A1.
PD	30-DEC-1998.
PF	18-JUN-1998; U12764.
PR	03-SEP-1997; US-057483.
PR	20-JUN-1997; US-050359.
PR	22-JUL-1997; US-053344.
PR	22-JUL-1997; US-053377.
PA	(HUMA-) HUMAN GENOME, SCI INC.
PA	(MEDT-) MEDIMUNE INC.
PI	Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,

RESULT	7
V22738/C	
ID	V22738 standard; DNA; 2287 BP.
AC	V22738;
DT	28-SEP-1998 (first entry)
DE	Babesia microti BMN1-8 antigen sequence.
KW	antigen; detection; Lyme disease; Vaccine; tick-borne disease;
OS	differentiation; Lyme disease; ehrlichiosis; ss.
DE	Babesia microti.
Key	location/Qualifiers
CDs	31.1806
FT	
FT	/*tag= a
FT	/product= antigen
PN	EP-834567-A2.
PD	08-APR-1998.
PF	01-OCT-1997; 117067.
PR	24-APR-1997; US-845258.
PR	01-OCT-1996; US-723142.
PA	(CORI-) CORIXA CORP.
PI	Houghton R, Lodes MJ, Reed SG, Sleath PR;
DR	WPI; 98-195465/18.
DR	P-P50B; W56287.
PT	Polypeptides comprising Babesia microti antigens and their
PT	immunogenic fragments or epitopes - and related nucleic acid,
PT	vectors, transformed cells and antibodies, useful for diagnosis of
PT	infection and in protective vaccines

PS Claim 8; Page 27-29; 113pp; English.
 CC The sequence is that encoding a polypeptide comprising at least
 CC one antigenic portion of a Babesia microti antigen. It can be used
 CC to diagnose B. microti infection by detecting specific antibodies
 CC in usual immunoassays. Infection can also be diagnosed using:
 CC (a) primers or probes derived from the coding sequence, in
 CC standard amplification or hybridisation tests, or (b) using
 CC antibodies to detect the corresponding antigen. It is also
 CC useful in vaccines to protect against infection, especially
 CC when formulated with an adjuvant. The new diagnostic methods
 CC allow rapid differentiation between B. microti infection and
 CC other tick-borne diseases (Lyme disease and ehrlichiosis) that
 CC have similar symptoms but require different treatments.
 CC Sequence 2287 BP; 981 A; 253 C; 250 G; 803 T;

Query Match 10.1%; Score 38.6; DB 1; Length 2287;
 Best Local Similarity 44.1%; Pred. No. 1.5; Mismatches 178; Indels 0; Gaps 0;
 Matches 143; Conservative 3;

OY 59 ATTGGTAATATGCAATTAATTAATTTTGGGCAATTCCTTACATGCTTGACAA 118
 DB 2133 ATTAGTAATGATATGATGATGATGATGATGATGATGATGATGATGATGAT 2074
 OY 119 GATTAAATGCTGTCGCAAAATTTGATTTTATTTGAGACTCTTATCAAAAGTAAT 178
 DB 2073 TATTAATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2014
 OY 179 GCTGCCAAGAGAGCTAGTAAGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 238
 DB 2013 TATGATATTTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1954
 OY 239 TCTAAAGATTTTGAATTCCTGGAATGACAAATTAATTAATTAATTAATTAATTAATTAAT 298
 DB 1953 TTAATTTGAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1894
 OY 239 TTAATGAGACCAAGCTTCTGATCTGATCTGATTAATTAATTAATTAATTAATTAATTAAT 358
 DB 1893 TTCTCAATTAATTAATTTTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1834
 OY 359 TGACATTAAGCTATATGTTTAA 382
 DB 1833 TTAATTAATTTTACTTCAATTA 1810

RESULT 8
 V74419
 ID V74419 standard; DNA: 4093 BP.

AC V74419; 16-MAR-1999 (first entry)
 DE Staphylococcus aureus contig SEQ ID #108.
 KW Computer readable medium; vaccine; S. aureus infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.

OS Staphylococcus aureus.
 FH Key Location/Qualifiers
 FT misc-feature 421. 480

FT /tag- a
 FT /note- "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"

FT misc-feature 2221. 2280
 FT /tag- b
 FT /note- "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"

FT misc-feature 4021. 4080
 FT /tag- c
 FT /note- "these bases represent a line of missing text in
 FT the sequence listing in the specification. They

are included to maintain the nucleotide numbering
 given in the specification for this DNA sequence"

FT EP-786519-A2.
 PN 30-JUL-1997;
 PF 07-JAN-1997; 100117
 PA 05-JAN-1996; US-009861.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
 PI Rosen CA;
 DR WPI: 97-374922/35.
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
 PT stored on computer readable medium and used in the production of
 PT anti-S. aureus vaccines
 PS Claim 1; Page 624-627; 3271pp; English.
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the S. aureus DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against S. aureus infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC S. aureus in a sample. S. aureus is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S. aureus DNA sequences contained on the
 CC computer readable medium.
 SQ Sequence 4093 BP; 1261 A; 757 C; 488 G; 1400 T;

Query Match 9.9%; Score 38; DB 1; Length 4093;
 Best Local Similarity 51.2%; Pred. No. 2.1;
 Matches 86; Conservative 1; Mismatches 81; Indels 0; Gaps 0;

OY 6 AATCAACAAAAAACAATCTTCATTCAGAAAAATATCTTAGGGCTGATATGTA 65
 DB 1633 AATATTAATAAATAATAGATACACGCTTCATCTATTAATATGATTAATGATGTC 1692
 OY 66 ATTATGCTCAATTAATTAATTTTGGGCAATTCCTTACATGCTGCAAGATTAA 125
 DB 1693 ACTCTCATCAATTTATTTTAAATACAGCTTTCATGATTAATTAAGCACGCTTAA 1752
 OY 126 ATGCTGTGCCAAAAATTTGATTTTATTTGAGAGACTCTTATCAAAA 173
 DB 1753 ATGTAAGTACGAAATCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1800

RESULT 9
 V01871/c
 ID V01871 standard; DNA: 2636 BP.

AC V01871; 20-APR-1998 (first entry)
 DE Human brain specific nucleosome assembly protein gene.
 KW Human; foetal brain cDNA library; GTP dissociation stimulating protein;
 KW brain specific nucleosome assembly protein; diagnosis; therapy;
 KW skeletal muscle specific ubiquitin conjugating enzyme; TMP-2; NPik;
 KW nel-related protein type 1; nel-related type 2; hereditary disease;
 KW cancer; ss.

OS Homo sapiens.

FH Key Location/Qualifiers
 FT CDS 266. 1786

FT /tag- a
 FT /product- "brain specific nucleosome assembly protein"
 EP-796913-A2.
 PN 24-SEP-1997;
 PF 19-MAR-1997; 104842.
 PR 05-MAR-1997; JP-069163.
 PR 19-MAR-1996; JP-063410.

PA (SAKA) OTSUKA PHARM CO LTD.
 PI Fujiwara T, Horie M, Watanabe T;
 DR WPI: 97-459830/43.
 DR P-PSDB: W37495.
 PT Novel human genes, e.g. brain-specific nucleosome assembly protein -
 useful for diagnosis or therapy of hereditary disease and cancer
 PS Claim 6: Page 60-64; 123pp; English.
 CC The present sequence encodes a brain specific nucleosome assembly
 protein isolated from a human foetal brain cDNA library. The nucleotide
 or amino acid sequences are useful for in-vitro diagnosis of hereditary
 diseases and cancer and for preparation of pharmaceuticals.
 CC Sequence 2636 BP; 814 A; 521 C; 599 G; 702 T;

Query Match 9.8%; Score 37.6; DB 1; Length 2636;
 Best Local Similarity 46.6%; Pred. No. 2.5; Mismatches 122; Indels 0; Gaps 0;

Matches 109; Conservative 3; Mismatches 122; Indels 0; Gaps 0;
 QY 55 TGATATTGTAATTTGTCATTTATTTATTTTGGGCGATTTCCCTTACATTTGCTTG 114
 DB 2039 TGCCTTAGATATTTGACACAGCTTACATTTTGAAGAAATTTACTACACAGCTTA 1980
 QY 115 ACAAGATTTAAATGCTGTCGCCAAATTTGTAATTTTGGAGACTTCTTATCAAAG 174
 DB 1979 GATGAAATTAACCTGACCTTAGACACTTTTGAACATTTTAAATATAGACTACCA 1920
 QY 175 TAATGCTGCCAAGAGATCTAAGAAATTTAGTCTCCCTCCTGTTGGAGTGTG 234
 DB 1919 GAAACCCAAACCTTGACATTAAGAAATTAAGTTCAGGTTAGGTTAGGTTTGA 1860
 QY 235 CTATCTAAAGATTTGATTTCTCGATGACATTTATTTTAACTTTGGTG 288
 DB 1859 CTTTTCAGGCGTGATGATCTGCTTCACTTCTTGAGATTTTAAAGTTCTTG 1806

RESULT 10

ID X19427 standard; DNA; 685 BP.
 AC X19427;
 DT 04-JUN-1999 (first entry)
 DE Nicotiana tabacum matrix attachment region pS211-1 DNA sequence.
 KW Nicotiana tabacum; matrix attachment region; MAR; tobacco;
 KW expression; transgene; structural gene; ss.
 OS Nicotiana tabacum.
 PN W0907865-21.
 PD 18-FEB-1999.
 PE 05-AUG-1998; U16344.
 PR 06-AUG-1997; US-066118.
 PA (UUNC-) UNIV NORTH CAROLINA STATE.
 PI Michalowski SM, Spiker SL;
 DR WPI: 99-167436/14.
 PT New Matrix Attachment Region DNA - useful for increasing expression
 of transgene product, especially structural genes in tobacco
 PS Claim 1; Fig 3; 66pp; English.
 CC The present sequence represents DNA encoding a Matrix Attachment Region
 (MAR) (which bind to the nuclear matrix) isolated from Nicotiana tabacum.
 CC A DNA construct containing MAR sequences can be used to increase
 expression of a transgene product. A new method identifies MARs by
 CC sequence alone, unlike prior art methods using associated genes. MARs
 CC increase expression of a transgene product.
 CC Sequence 685 BP; 243 A; 107 C; 87 G; 248 T;

Query Match 9.7%; Score 37.2; DB 1; Length 685;
 Best Local Similarity 49.4%; Pred. No. 2.6;
 Matches 87; Conservative 3; Mismatches 86; Indels 0; Gaps 0;

QY 31 ATTCAGAAAATTTACTAGGAGCATGATTTGTAATATGCTCATTTATTTTCTT 90
 DB 366 ATTTAAAAAATTAAGTGGGCGCTGCAATTTTATTTATTTATTTATTTT 425
 QY 91 GGGGCAATTCCTTACATTTGCTTGACAGATTTAAATGCTGCGCAAAATTTGTAATT 150

DB 426 AGGAGATCCCTCCCTTATTTATGAAATACCTTTAATGACTACATCTTATTTACT 485
 QY 151 TATTTGAGACTCTTTTCAAAAGTATGCTGCCAAAGAGCTTAAGGAATAGT 206
 DB 486 AAGTTGTCTATATTAATGATGCAATCTCTCATATCAATAAATATACATTTAAT 541

RESULT 11

Continuation (8 of 17) of V21209 from base 700001 (Methanococcus jannaschii circular
 V21209.07
 WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

WP	Fragment Name	Begin	End
WP	V21209.00	1	110000
WP	V21209.01	100001	210000
WP	V21209.02	200001	310000
WP	V21209.03	300001	410000
WP	V21209.04	400001	510000
WP	V21209.05	500001	610000
WP	V21209.06	600001	710000
WP	V21209.07	700001	810000
WP	V21209.08	800001	910000
WP	V21209.09	900001	1010000
WP	V21209.10	1000001	1110000
WP	V21209.11	1100001	1210000
WP	V21209.12	1200001	1310000
WP	V21209.13	1300001	1410000
WP	V21209.14	1400001	1510000
WP	V21209.15	1500001	1610000
WP	V21209.16	1600001	1664976

Query Match 9.6%; Score 36.8; DB 1; Length 110000;

Best Local Similarity 45.2%; Pred. No. 5.7; Mismatches 145; Indels 0; Gaps 0;

Matches 122; Conservative 3; Mismatches 145; Indels 0; Gaps 0;

QY 18 AAAACATTTCTTCATTCAGAAAATTTATCTTAGGAGCTGATTTGTAATTTAGTCAAT 77
 DB 64033 ATAGATATTTTTCACACTTTTATTTTATTTTTCGATTAATATTAATATTTTAAAT 64092
 QY 78 TTAATWRTTGTGGGCGATTTCTTACATTTGCTTGACAAAGATTAATGCTGTGCCA 137
 DB 64093 ATTTTCGACTTAAATTTTAAATGATTTTTCGCAAAAAGAAACAAATTTCTGT 64152
 QY 138 AAATTTGATTTTATTTGAGACTCTTATCAAAAGTATGCTGCCAAAGAGCTTA 197
 DB 64153 AAATTTTATGATTTTATGAAAAAATGAAAAATTTCTTAATTAATAAATAAAT 64212
 QY 198 GAATTTAGTGTGTCCTCCTGCTGTTGAGAGTGTCTAATTTCAAAAGATTTGATTTG 257
 DB 64213 TAGAAATTAAGTTAAATTTACATTTTATTTGATTTTAAATTTTACCTTATTTAT 64272
 QY 258 CTGGAATGACATTTATTTTAACTTTGGT 287
 DB 64273 TCCTTTGAACATTTATTTTCAATTTTGT 64302

RESULT 12

Q27886/C
 ID Q27886 standard; DNA; 2418 BP.
 AC Q27886;
 DT 04-FEB-1993 (first entry)
 DE P.falciparum GAP130h.
 KW Polymerase chain reaction; glycoprotein binding protein; inverse PCR;
 KW Plasmodium falciparum; malaria; ss.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT exon 767..955
 FT exon /tag- a
 FT exon /number- 1
 FT exon 1111..2202
 FT exon /tag- b
 FT exon /number- 2
 FT repeat_region 1249..2202

```

FT      EP-499834-A.      /*tag= c
PN      26-AUG-1992.
PD      27-JAN-1992: 101271.
PR      21-FEB-1991: DE-105348.
PA      (BEHM) BEHRINGWERKE AG.
PI      Hundt E, Knapp B, Kupper H, Nolte D, Kuepper H;
PT      WPI: 92-286009/35.
PS      DNA coding protein GBP 130 h and proteins produced - used for
PS      vaccines to control malaria
PS      Claim 1: Page 12-14; 17pp; German.
CC      Glucoprotein binding protein 130h is a P.falciparum blood stage
CC      antigen 69k homologous with the known GBP130. The GBP130h gene and
CC      its homologues can be used in vaccines to protect against malaria
CC      caused by P.falciparum. The coding sequence was isolated using PCR
CC      techniques on a genomic P.falciparum FCBR strain library. The
CC      GBP130h gene was found to be highly conserved between different
CC      strains but is distinct from the GBP130 gene.
CC      See Q27878-Q27885 and Q33432.
SQ      Sequence 2418 BP; 1047 A; 303 C; 292 G; 776 T;

Query Match
Best Local Similarity 9.4%; Score 36; DB 1; Length 2418;
Matches 159; Conservative 4; Mismatches 210; Indels 0; Gaps 0;

OY      1 AAACAACAACAACAACAACAACAATCTTCATCAGAAAATATATCTTAGGAGCTGATAT 60
DB      638 AAAAAAAAAAAAAAAAAAACAATACATATAATTTTATTTATTTATTTATTTT 579
OY      61 TCGTAATTAATGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 120
DB      578 TTTTATTTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 519
OY      121 TTAATAATGCTGCGCAAAATTTGATTTTATTTGAGACTCTTATCAAAAGTATGC 180
DB      518 AATTTTTCATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 459
OY      181 TSCCAAGAGAGCTCAAGAAATAGTAGTGTCCCTCACTTTGAGAGTGTCTATTC 240
DB      458 TATGTAGTATACAGAGAGCTAGAAATATATATATATATATATATATATATATATAT 399
OY      241 TAAAGATTTGATTTCTCGGAATGACATTAATTAATTAATTAATTAATTAATTAATTA 300
DB      398 AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 339
OY      301 AATAGGACCAAGCTCACTGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 360
DB      338 TTTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 279
OY      361 ACCATTAGCTAT 373
DB      278 TATTAATAAATAT 266

RESULT 13
T89990/C
ID      T89990 standard; DNA; 3607 BP.
AC      T89990;
DT      11-MAY-1998 (first entry)
DE      Cryptosporidium parvum oocyst heat shock protein HSP70 gene.
KW      Cryptosporidium; detection; assay; protozoan; heat shock protein;
KW      HSP70; gene amplification; PCR; ds.
OS      Cryptosporidium parvum.
FH      Key
FT      CDS
FT      1272..3296
FT      Location/Qualifiers
FT      /tag= a
FT      /transl_except= (pos:1536..1538, aa:His)
FT      /transl_except= (pos:1578..1580, aa:Leu)
FT      /transl_except= (pos:1659..1661, aa:Lys)
FT      /transl_except= (pos:2034..2036, aa:Lys)
FT      /transl_except= (pos:2265..2267, aa:Leu)
FT      /transl_except= (pos:2931..2933, aa:Lys)

```

```

FT      MO9742349-A1.
PN      13-NOV-1997.
PR      08-MAY-1997; U07972.
PF      09-MAY-1996; US-647351.
PA      (METR-) METROPOLITAN WATER DISTRICT SOUTHERN CALIFORNIA.
PI      De Leon R, Rochelle PA;
PT      WPI: 97-55899/51.
PS      P-ESDB; W31345.
PS      Amplifying and detecting viable and infective protozoan oocysts -
PS      by monitoring presence or absence of heat shock protein genes in
PS      their transcripts
PS      Claim 12: Page 23-26; 44pp; English.
CC      This genomic DNA sequence comprises the heat shock protein HSP70
CC      gene of Cryptosporidium parvum oocysts. The invention provides
CC      processes and kits for detecting encysted forms of protozoa,
CC      particularly Cryptosporidium (especially Cryptosporidium muris,
CC      Cryptosporidium baileyi, Cryptosporidium wrairi and C. parvum)
CC      and Giardia, that are viable and infectious by the enzymatic
CC      amplification (e.g. by PCR) of a target gene sequence such as HSP70.
CC      A preferred primer pair (see T89991-92) can amplify DNA or RNA
CC      from many different Cryptosporidium species, and can be used with a
CC      probe (see T89997) that confirms the identity of the amplification
CC      product. A most preferred primer pair (see T89993-94) is specific
CC      to C. parvum and can also be used with a probe (see T89998) to
CC      confirm sequence identity. In addition, both Cryptosporidium and
CC      Giardia can be detected simultaneously using a multiplex
CC      amplification reaction. The methods exploit the speed, sensitivity
CC      and specificity associated with an amplification procedure, and
CC      enable pathogenic forms of protozoa present in low copy numbers to
CC      be identified and distinguished from morphologically similar, but
CC      non-pathogenic protozoa.
SQ      Sequence 3607 BP; 1265 A; 514 C; 716 G; 1112 T;

Query Match
Best Local Similarity 9.4%; Score 36; DB 1; Length 3607;
Matches 87; Conservative 4; Mismatches 89; Indels 0; Gaps 0;

OY      1 AAACAACAACAACAACAACAACAATCTTCATCAGAAAATATATCTTAGGAGCTGATAT 60
DB      221 AAAAAAAAAAAAAAAAAAATACACTTTATTAATTAATTAATTAATTAATTAATTAATTA 162
OY      61 TCGTAATTAATGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 120
DB      161 TGAATCTCCTCATCAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 102
OY      121 TTAATAATGCTGCGCAAAATTTGATTTTATTTGAGACTCTTATCAAAAGTATGC 180
DB      101 TGAACATATGCTTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 42

RESULT 14
V21209_05
Continuation (6 of 17) of V21209 from base 500001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209
WP Fragment Name Begin End
WP V21209_00 1 110000
WP V21209_01 100001 210000
WP V21209_02 210001 310000
WP V21209_03 310001 410000
WP V21209_04 410001 510000
WP V21209_05 510001 610000
WP V21209_06 610001 710000
WP V21209_07 710001 810000
WP V21209_08 810001 910000
WP V21209_09 910001 1010000
WP V21209_10 1010001 1110000
WP V21209_11 1110001 1210000
WP V21209_12 1210001 1310000
WP V21209_13 1310001 1410000

```


No.	Score	Match Length	DB	ID	Description
C 1	37.6	9.8	2636	4	US-08-8270-170A-21
C 2	36	9.4	3607	3	US-08-8677-351B-1
C 3	34.8	9.1	5852	1	US-07-867-106-2
C 4	34.6	9.0	2334	2	US-08-062-632-4
C 5	34.2	8.9	2749	1	US-08-344-695-1
C 6	34	8.9	5852	1	US-07-867-106-2
C 7	33.6	8.8	8920	4	US-08-446-855A-1
C 8	31.6	8.3	2074	2	US-08-414-685-1
C 9	31.6	8.3	1422	3	US-08-319-704-5
C 10	31.4	8.2	2636	3	US-08-253-785-1
C 11	31.2	8.1	901	3	US-08-832-883-65
C 12	31.2	8.1	901	4	US-08-832-877-65
C 13	30.8	8.0	7101	3	US-08-480-604A-9
C 14	30.6	8.0	414	1	US-08-377-687-8
C 15	30.6	8.0	8878	1	US-08-206-176-3
C 16	30.6	8.0	414	3	US-08-777-192-48
C 17	30.6	8.0	1779	5	PCT-US96-05320X-831
C 18	30.4	7.9	1939	1	US-07-715-751B-2
C 19	30.4	7.9	1511	1	US-07-991-867B-8
C 20	30.4	7.9	660	1	US-07-991-867B-32
C 21	30.4	7.9	3661	1	US-08-105-483-221
C 22	30.4	7.9	3661	2	US-08-303-124-7
C 23	30.4	7.9	3661	2	US-08-204-729-7
C 24	30.4	7.9	2762	2	US-08-198-446B-12
C 25	30.4	7.9	1511	2	US-08-107-755A-8
C 26	30.4	7.9	660	2	US-08-417-755A-32
C 27	30.4	7.9	3660	3	US-08-475-063-35
C 28	30.4	7.9	3660	3	US-08-207-792-25
C 29	30.4	7.9	3661	3	US-08-709-209-221
C 30	30.4	7.9	3661	3	US-08-488-101-221
C 31	30.4	7.9	2762	4	US-08-870-693-12
C 32	30.4	7.9	4951	4	PCT-US95-06119-5
C 33	30.2	7.9	1560	3	US-08-356-180-1
C 34	30.2	7.9	2213	3	US-08-592-936B-11
C 35	30.2	7.9	2145	3	US-08-592-936B-16
C 36	30.2	7.9	2917	3	US-08-592-936B-20
C 37	30.2	7.9	1336	3	US-08-552-936B-22
C 1	37.6	9.8	2636	4	Sequence 21, App
C 2	36	9.4	3607	3	Sequence 1, App1
C 3	34.8	9.1	5852	1	Sequence 2, App1
C 4	34.6	9.0	2334	2	Sequence 4, App1
C 5	34.2	8.9	2749	1	Sequence 1, App1
C 6	34	8.9	5852	1	Sequence 2, App1
C 7	33.6	8.8	8920	4	Sequence 1, App1
C 8	31.6	8.3	2074	2	Sequence 1, App1
C 9	31.6	8.3	1422	3	Sequence 5, App1
C 10	31.4	8.2	2636	3	Sequence 1, App1
C 11	31.2	8.1	901	3	Sequence 65, App1
C 12	31.2	8.1	901	4	Sequence 65, App1
C 13	30.8	8.0	7101	3	Sequence 9, App1
C 14	30.6	8.0	414	1	Sequence 48, App1
C 15	30.6	8.0	8878	1	Sequence 3, App1
C 16	30.6	8.0	414	3	Sequence 48, App1
C 17	30.6	8.0	1779	5	Sequence 48, App1
C 18	30.4	7.9	1939	1	Sequence 81, App1
C 19	30.4	7.9	1511	1	Sequence 2, App1
C 20	30.4	7.9	660	1	Sequence 8, App1
C 21	30.4	7.9	3661	1	Sequence 32, App1
C 22	30.4	7.9	3661	2	Sequence 21, App1
C 23	30.4	7.9	3661	2	Sequence 7, App1
C 24	30.4	7.9	2762	2	Sequence 12, App1
C 25	30.4	7.9	1511	2	Sequence 8, App1
C 26	30.4	7.9	660	2	Sequence 32, App1
C 27	30.4	7.9	3660	3	Sequence 25, App1
C 28	30.4	7.9	3660	3	Sequence 25, App1
C 29	30.4	7.9	3661	3	Sequence 221, App1
C 30	30.4	7.9	3661	3	Sequence 221, App1
C 31	30.4	7.9	2762	4	Sequence 221, App1
C 32	30.4	7.9	4951	4	Sequence 12, App1
C 33	30.2	7.9	1560	3	Sequence 1, App1
C 34	30.2	7.9	2213	3	Sequence 12, App1
C 35	30.2	7.9	2145	3	Sequence 16, App1
C 36	30.2	7.9	2917	3	Sequence 20, App1
C 37	30.2	7.9	1336	3	Sequence 22, App1

38	30.2	7.9	2213	3	US-08-788-928A-1	Sequence 1, Appl1
39	30.2	7.9	1356	3	US-08-788-928A-2	Sequence 2, Appl1
40	30.2	7.9	259	3	US-08-788-928A-4	Sequence 4, Appl1
41	30.2	7.9	246240	4	US-08-724-394A-20	Sequence 20, Appl1
42	30.2	7.9	246240	4	US-08-724-394A-21	Sequence 21, Appl1
43	30.2	7.9	246240	4	US-08-724-394A-22	Sequence 22, Appl1
44	30	7.8	731	3	US-08-451-405A-2	Sequence 2, Appl1
45	30	7.8	246240	4	US-08-724-394A-22	Sequence 22, Appl1

ALIGNMENTS

RESULT 1

US-08-820-170A-21/c
Sequence 21, Application US/08820170A
Patent No. 5831058
GENERAL INFORMATION:
APPLICANT: TSUTOMU, FUJIWARA
APPLICANT: TAKESHI, WATANABE
APPLICANT: MASATO, HORIE
APPLICANT: TOYOMASA, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sugitue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
City: Washington
STATE: D.C.
COUNTRY: United States
zip: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.3GD
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/820,170A

CLASSIFICATION: 536
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 293-7060
 TELEFAX: (202) 293-7860
 TELEX: 6491103
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2636 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA(genomic)
 HYPOTHEetical: NO
 ANTI-SENSE: NO
 IMMEDIATE SOURCE:
 LIBRARY: Human fetal brain cDNA library
 CLONE: GEN-078D05
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 266..1783
 OS-08-920-170A-21

Query Match 9.88; Score 37.6; DB 4; Length 2636;

Matches 109; Conservative 3; Mismatches 122; Indels 0; Gaps 0;

55 TGATATTGGTAATTATGGTCAATTAAWRTRTKTGGGCATTCCCTTACATTGCTTG 114

Db 2039 TGCTTAGATAATGACACAGCTTACATATTTGAGAATATTATACTACACAGCCTA 1980

QY 115 ACAGATTAAATGTCGTGCCAAATTTGTATTATTGAGACTCTTATCAAAG 174

Db 1979 GATAGATAACTGGCACTTTTAGACTATTTTAAATATAGACTACCA 1920

OY 175 TATGCTGCCAAGAGAGCTAAGAAATAGTAGTGTCCCTACCTTTGAGAGTGT 234
DB 1919 GAACACCAAAAACCTGACTATAGAAATAGTGTACAGTACAGGCTTTTA 1860
OY 225 CTTATCTAAAGATTTTGATTTCTCGGAATGACAAATTATTTTAACTTTGGTG 288
DB 1859 CTTTTCCTCAAGCGCTGATGAACTGCTTCACTTGTGAGATTTTAAAGATTCGTG 1806

RESULT 2
US-08-647-351B-1/c

; Sequence 1, Application US/08647351B
; Patent No. 5770368
; GENERAL INFORMATION:
; APPLICANT: De Leon, Ricardo
; APPLICANT: Rochelle, Paul
; TITLE OF INVENTION: Cryptosporidium Detection Method
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Max
; STREET: 225 S. Lake Avenue, 9th Floor
; CITY: Pasadena
; STATE: California
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows version 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/647, 351B
; FILING DATE: May 9, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Farah, David A.
; REGISTRATION NUMBER: 38,134
; REFERENCE/DOCKET NUMBER: 11364
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 796-4000
; TELEFAX: (626) 795-6321
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3607 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double stranded
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-647-351B-1

Query Match 9.4%; Score 36; DB 3; Length 3607;
Best Local Similarity 48.3%; Pred. No. 0.72;
Matches 87; Conservative 4; Mismatches 89; Indels 0; Gaps 0;

OY 1 AAACAACAACAACAACAACAACAATCTTCATTCAGAAAAATTAATCTAGGAGCTGAT 60
DB 221 AATAAATAAATAAATAAATAAATTAATTAATAAATAAATAAATAAATAAATAA 162
OY 61 TCGTAATTAATGCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 120
DB 161 TCGAATCTCTCATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 102
OY 121 TTAATAATGCTGCGCAAAATTTTGAATTTTATTTGAGACCTCTTAATAAAGTAA 180
DB 101 TGAACCTAAATGCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 42

RESULT 3
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:

; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526-1s
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeoney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3439
; TELEFAX: 215-568-3100
; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; US-07-867-106-2

Query Match 9.1%; Score 34.8; DB 1; Length 5852;
Best Local Similarity 48.9%; Pred. No. 1.6;
Matches 87; Conservative 2; Mismatches 89; Indels 0; Gaps 0;

OY 1 AAACAACAACAACAACAACAACAATCTTCATTCAGAAAAATTAATCTAGGAGCTGAT 60
DB 2022 AAAAAAACAACAACAACAACAACCTCATTAATAATTAATTAATTAATTAATTA 2081
OY 61 TCGTAATTAATGCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 120
DB 2082 TTTTATTTTAAATAAATAAATTTTATTCATCTAATTAATTAATTAATTAATTA 2141
OY 121 TTAATAATGCTGCGCAAAATTTTGAATTTTATTTGAGACCTCTTAATAAAGTAA 178
DB 2142 TTGCAATTAATTAATAAATTTTATTAATTAATTAATTAATTAATTAATTAAT 2199

RESULT 4
US-08-062-632-4/c
; Sequence 4, Application US/08062632
; Patent No. 5712090
; GENERAL INFORMATION:
; APPLICANT: Artushin, Sergey
; APPLICANT: Stipkovits, Laslo

APPLICANT: Munion, F. Chris
TITLE OF INVENTION: PCR-Based Assay for Mycoplasma
TITLE OF INVENTION: Hypneumonidae
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dickstein, Shapiro and Morin
STREET: 2101 L. St. NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 22037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/062,632
FILING DATE: 18-MAY-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brady Jr., James W.
REGISTRATION NUMBER: 32,115
REFERENCE/DOCKET NUMBER: 18900.018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)785-9700
TELEFAX: (202)887-0689
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2334 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-062-632-4

Query Match 9.0%; Score 34.6; DB 2; Length 2334;
Best Local Similarity 48.6%; Pred. No. 1.4; Matches 85; Conservative 3; Mismatches 87; Indels 0; Gaps 0;

QY 1 AAAAAACAAACAAAAAATCTTCATTCAGAAAAATATATCTAGGACTGATAT 60
DB 496 AAAAAAGAGAGAGATTTATATAACATGAAATATGATATATTCGCAATAT 437
QY 61 TGGTAATATGTCATTTAATWRTTGTGGGCGACTTCCTTACATTTGCTTGACAAGA 120
DB 436 CAGAAATAAAGCTCCTCTTACTTAATCAAAATTAATTTACTATAGAAATTTATATTA 377
QY 121 TTAATATGTCGCGCAAAATTTTGTATTTTATTTGAGACTCTTATCAAAAGT 175
DB 376 TTAGATGTTGCTCTCAAAATTCAGCACTTTTATATTAATCAAAATTTAAAT 322

RESULT 5
US-08-344-695-1/c
Sequence 1, Application US/08344695
Patent No. 5614398
GENERAL INFORMATION:
APPLICANT: O'BROCHTA, DAVID
APPLICANT: WARREN, WILLIAM
APPLICANT: ATKINSON, PETER
TITLE OF INVENTION: A GENE TRANSFER SYSTEM FOR INSECTS
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/344,695
FILING DATE: 18-NOV-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Keiber, Steven B.
REGISTRATION NUMBER: 30,073
REFERENCE/DOCKET NUMBER: 2747-058-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2749 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 450..2285
US-08-344-695-1

Query Match 8.9%; Score 34.2; DB 1; Length 2749;
Best Local Similarity 49.2%; Pred. No. 1.9; Matches 87; Conservative 1; Mismatches 89; Indels 0; Gaps 0;

QY 14 AAAAAAACAAATCTTCATTCAGAAAAATATATCTAGGACTGATATGTAATATGCT 73
DB 1147 AAAAAATGCTTAAGCTCTTAATATATTTCTGCTGAGCTTTCAGAAATCTAG 1088
QY 74 CAATTAATWRTTGTGGGCAATTCCTTACATTTGCTTGACAAATGATTAATGCTCT 133
DB 1087 GACTTTAACTTAATATGATCTGCACTTCAATTTGTTTCAATGAGTAGTAACCTTACT 1028
QY 134 GCAAAATTTGATTTTATTTGAGACTCTTATCAAAAGTAATGCTGCAAGAAGA 190
DB 1027 CCAAAATTCGCTTTATATATATATATGCGTCCAAATCTATGCTGCTATGCA 971

RESULT 6
US-07-867-106-2/c
Sequence 2, Application US/07867106
Patent No. 5389526
GENERAL INFORMATION:
APPLICANT: Slade, Martin B
APPLICANT: Chang, Andy C M
APPLICANT: Williams, Keith L
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625
PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU PJ 7187
APPLICATION NUMBER: PCT/AU90/00530
FILING DATE: 02-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Feeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
US-07-867-106-2

Query Match 8.9%; Score 34; DB 1; Length 5852;
Best Local Similarity 48.8%; Pred. No. 2.4;
Matches 82; Conservative 3; Mismatches 83; Indels 0; Gaps 0;

Oy 1 AAACAACAAACAAACAAACAAATTCCTTCAGAAAAATTATCTAGGACGTGATAT 60
Db 5626 AAAAAAAAAAAAAAAAAAAAAAAAAAATGATAATTTGCAATTAATAAATAA 5567
Oy 61 TGGTAATTATGCTCAATTAATWRTTRTGGGCAATTCCTTACATGCTTGACAGA 120
Db 5566 AATGTAAGGGGTTTTTTTAAATATGATGATGATTTTAAATCAATTCAGAGA 5507
Oy 121 TTAATATGCTGCGCAAAATTTTGTATTTTATTTGAGACCTCTAT 168
Db 5506 TTAATAAACTTAAACAAATAAACAATATTTGATTTTATTTTATTTT 5459

RESULT 7
US-08-446-855A-1
Sequence 1, Application US/0844685A
Patent No. 5849573
GENERAL INFORMATION:
APPLICANT: Stewart, Thomas S
APPLICANT: Flores, Maria V
APPLICANT: O'Sullivan, William J
TITLE OF INVENTION: Nucleotide sequence encoding cardamomyl
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 1100 No. 5849573th Glebe Road, 8th Floor
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,855A
FILING DATE: 06-Jul-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mitchard, Leonard C

REGISTRATION NUMBER: 29,009
REFERENCE/DOCKET NUMBER: 47-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic
US-08-446-855A-1

Query Match 8.8%; Score 33.6; DB 4; Length 8920;
Best Local Similarity 59.3%; Pred. No. 3.3;
Matches 51; Conservative 3; Mismatches 32; Indels 0; Gaps 0;

Oy 1 AAACAACAAACAAACAAACAAATTCCTTCAGAAAAATTATCTAGGACGTGATAT 60
Db 633 AAAAAAAAAAAAAAAAAAAAAAAAAAATTTACATATGAAAAATGAACTGTATGTAAT 692
Oy 61 TGGTAATTATGCTCAATTAATWRTTR 86
Db 693 TTATAATATTTTAAACATAAATA 718

RESULT 8
US-08-414-685-1
Sequence 1, Application US/08414685
Patent No. 5667989
GENERAL INFORMATION:
APPLICANT: McCullough, John
APPLICANT: Baymiller, Judy
TITLE OF INVENTION: Fungal Cell Wall Protein CLY4
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burton Rodney
STREET: P.O. Box 4000
CITY: Princeton
STATE: New Jersey
COUNTRY: USA
ZIP: 08543-4000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/414,685
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Gaul, Timothy J.
REGISTRATION NUMBER: 33,111
REFERENCE/DOCKET NUMBER: DC33
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 252-5901
TELEFAX: (609) 252-4526
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2074 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 334..1193
US-08-414-685-1

RESULT 11
US-08-832-883-65/c
; Sequence 65, Application US/08832883
; Patent No. 5807681
; GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; APPLICANT: Balidi, Alphonso
; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS
; TITLE OF INVENTION: OF CANCER
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
; STREET: Suite 1800 Two Penn Center Plaza
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/832,883
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-13 US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 901 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-883-65

Query Match 8.1%; Score 31.2; DB 3; Length 901;
Best Local Similarity 47.1%; Pred. No. 8.2; Mismatches 122; Indels 2; Gaps 1;
Matches 114; Conservative 4;
QY 20 AACATCTCTCATTCAGAAAATTAATCTTGGGACTGATATGTAATTTGGTCAATT 79
DB 774 AATACTTAACCTAAGAAAATTAATCTTATAGCAAACTTAAGTGTGTTGGCCA 715
QY 80 AATWRTTRTKGGGGCATTTCTTACAT--TGCTTGACAAAGATTAATATGCTGGCCA 137
DB 714 GTATATATGTAAGCAATTTTATACATTTATGTCATGTCCTTACAAACCCCATAT 655
QY 138 AATTTTGTATTTATTTGAGACTTCTTATCAAAAGTATGCTGCGCAAGAGCTAA 197
DB 654 GCTAGGAACTAGTTATTTCCATTTTATATAGAGAAATGAGTACAGAGAAATTTA 595
QY 198 GGAATTAGTAGTGTCCCTCACTTGTGAGTGTGCTATTTCTAAAGATTTGATTC 257
DB 594 ATGATTTGACAGGTTTACACACTGTTAAGTACTGGGAAATTTGAACCCATGTAAGCTGC 535
QY 258 CT 259
DB 534 CT 533

RESULT 12
US-08-832-877-65/c
; Sequence 65, Application US/08832877
; Patent No. 5840506

GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS OF
; TITLE OF INVENTION: CANCER
; NUMBER OF SEQUENCES: 116
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
; STREET: Suite 1800 Two Penn Center Plaza
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/832,877
; FILING DATE:
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-13 US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 901 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-877-65

Query Match 8.1%; Score 31.2; DB 4; Length 901;
Best Local Similarity 47.1%; Pred. No. 8.2; Mismatches 122; Indels 2; Gaps 1;
Matches 114; Conservative 4;
QY 20 AACATCTCTCATTCAGAAAATTAATCTTGGGACTGATATGTAATTTGGTCAATT 79
DB 774 AATACTTAACCTAAGAAAATTAATCTTATAGCAAACTTAAGTGTGTTGGCCA 715
QY 80 AATWRTTRTKGGGGCATTTCTTACAT--TGCTTGACAAAGATTAATATGCTGGCCA 137
DB 714 GTATATATGTAAGCAATTTTATACATTTATGTCATGTCCTTACAAACCCCATAT 655
QY 138 AATTTTGTATTTATTTGAGACTTCTTATCAAAAGTATGCTGCGCAAGAGCTAA 197
DB 654 GCTAGGAACTAGTTATTTCCATTTTATATAGAGAAATGAGTACAGAGAAATTTA 595
QY 198 GGAATTAGTAGTGTCCCTCACTTGTGAGTGTGCTATTTCTAAAGATTTGATTC 257
DB 594 ATGATTTGACAGGTTTACACACTGTTAAGTACTGGGAAATTTGAACCCATGTAAGCTGC 535
QY 258 CT 259
DB 534 CT 533

RESULT 13
US-08-480-604A-9
; Sequence 9, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.

;; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
;; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
;; NUMBER OF SEQUENCES: 32
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: MEDLEN & CARROLL, LLP
;; STREET: 220 MONTGOMERY STREET, SUITE 2200
;; CITY: SAN FRANCISCO
;; STATE: CALIFORNIA
;; COUNTRY: UNITED STATES OF AMERICA
;; ZIP: 94104
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: IBM PC compatible
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/480,604A
;; FILING DATE: 07-JUN-1995
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/422,711
;; FILING DATE: 14-APR-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/405,496
;; FILING DATE: 16-MAR-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/329,154
;; FILING DATE: 25-OCT-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/161,907
;; FILING DATE: 02-DEC-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/985,321
;; FILING DATE: 04-DEC-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/429,791
;; FILING DATE: 31-OCT-1989
;; ATTORNEY/AGENT INFORMATION:
;; NAME: INGOLTA, DIANE E.
;; REGISTRATION NUMBER: 40,027
;; REFERENCE/DOCKET NUMBER: OPHD-01763
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 705-8410
;; TELEFAX: (415) 397-8338
;; INFORMATION FOR SEQ ID NO: 9:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 7101 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..7098
;; US-08-480-604A-9

Query Match 8.0%; Score 30.8; DB 3; Length 7101;
Best Local Similarity 51.8%; Pred. No. 15;
Matches 59; Conservative 4; Mismatches 51; Indels 0; Gaps 0;

QY 8 ACAACAAAAAACAATCTTCATTCGAAAAATATCTTAGGACGCGATATGGTAAT 67
DB 4227 AGAAGGATTAATCAATTAAGAGTTGATTATATCTAAATCATATAAATTACTAT 4286

QY 68 TATGTCATTTAATWRTTRTKTGCGCATTCCTTACATTTGCTTGACAGAT 121
DB 4287 TTCGCGCAATTAATAAATATGATGTTAATTCAAATCATATTCACAGAAAT 4340

RESULT 14
US-08-377-687-48/C
; Sequence 48, Application US/08377687

;; Patent No. 5538525
;; GENERAL INFORMATION:
;; APPLICANT: BROEKERT, WILLEM F.
;; APPLICANT: CAMMUE, BRUNO P.A.
;; APPLICANT: OSBORN, ROBERT W.
;; APPLICANT: REES, SARAH B.
;; APPLICANT: TERRAS, FRANKY R.G.
;; APPLICANT: VANDERLEYDEN, JOZEF
;; TITLE OF INVENTION: BIOCIDAL PROTEINS
;; NUMBER OF SEQUENCES: 59
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
;; STREET: 1100 NEW YORK AVENUE, N.W.
;; CITY: WASHINGTON
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20005
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: IBM PC compatible
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/377,687
;; FILING DATE:
;; CLASSIFICATION: 800
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/002,480
;; FILING DATE: 04-JAN-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KOKULIS, PAUL N.
;; REGISTRATION NUMBER: 16,773
;; REFERENCE/DOCKET NUMBER: 99042/SEE.36525/US/A
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-822-0944
;; TELEFAX: 202-861-3000
;; INFORMATION FOR SEQ ID NO: 48:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 414 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: both
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 16..255
;; US-08-377-687-48

Query Match 8.0%; Score 30.6; DB 1; Length 414;
Best Local Similarity 46.9%; Pred. No. 9.8;
Matches 84; Conservative 3; Mismatches 92; Indels 0; Gaps 0;

QY 20 AACATCTTCATTCACAAAAAATATCTTAGGACGCGATATGGTAATGTCATTT 79
DB 391 AATAAACCTTATTTGATTAACGACCAACATTAATCTGTACATCTCTTAATC 332

QY 80 AATWRTTRTKTGCGCATTCCTTACATTTGCTTGACAGATTAATGCTGTGCATA 139
DB 331 ACGCATGAGATGACACGACGCTATTTGTAATATACATAAATCACTTCACCAAG 272

QY 140 ATTTGATTTTATTTTGAGACCTTCTTATCAAAAGTAATCTGCCAAAGAGCTTAAG 198
DB 271 AGTTGCGAATTAATTAACAAGGAAGTACGATACACTGTGAGCTGGAGACATAG 213

RESULT 15
US-08-206-176-3
; Sequence 3, Application US/08206176
; Patent No. 5639940
; GENERAL INFORMATION:
; APPLICANT: Garner, Ian
; APPLICANT: Dalrymple, Michael A

```

1  APPLICANT: Prunkard, Donna E
2  APPLICANT: Foster, Donald C
3  TITLE OF INVENTION: Production of fibrinogen in Transgenic
4  TITLE OF INVENTION: Animals
5  NUMBER OF SEQUENCES: 27
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE: ZymoGenetics, Inc.
8  STREET: 4225 Roosevelt Way, N.E.
9  CITY: Seattle
10 STATE: WA
11 COUNTRY: USA
12 ZIP: 98105
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: Patent Release #1.0, Version #1.25
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/206,176
20 FILING DATE:
21 CLASSIFICATION: 435
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Parker, Gary E
24 REGISTRATION NUMBER: 31-648
25 REFERENCE/DOCKET NUMBER: 93-15
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: 206-547-8080 ext 322
28 TELEFAX: 206-548-2329
29 INFORMATION FOR SEQ ID NO: 3:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 8878 base pairs
32 TYPE: nucleic acid
33 STRANDEDNESS: double
34 TOPOLOGY: linear
35 MOLECULE TYPE: DNA (genomic)
36 IMMEDIATE SOURCE:
37 CLONE: human fibrinogen B-beta chain
38 FEATURE:
39 NAME/KEY: misc_RNA
40 LOCATION: 1..469
41 FEATURE:
42 NAME/KEY: exon
43 LOCATION: 470..583
44 FEATURE:
45 NAME/KEY: intron
46 LOCATION: 584..3257
47 FEATURE:
48 NAME/KEY: exon
49 LOCATION: 3258..3449
50 FEATURE:
51 NAME/KEY: intron
52 LOCATION: 3450..3938
53 FEATURE:
54 NAME/KEY: exon
55 LOCATION: 3939..4122
56 FEATURE:
57 NAME/KEY: intron
58 LOCATION: 4123..5042
59 FEATURE:
60 NAME/KEY: exon
61 LOCATION: 5043..5270
62 FEATURE:
63 NAME/KEY: intron
64 LOCATION: 5271..5830
65 FEATURE:
66 NAME/KEY: exon
67 LOCATION: 5831..5944
68 FEATURE:
69 NAME/KEY: intron
70 LOCATION: 5945..6632
71 FEATURE:
72 NAME/KEY: exon
73 LOCATION: 6633..6758

```

```

? FEATURE:
? NAME/KEY: intron
? LOCATION: 6759..6966
? FEATURE:
? NAME/KEY: exon
? LOCATION: 6967..7252
? FEATURE:
? NAME/KEY: intron
? LOCATION: 7253..7870
? FEATURE:
? NAME/KEY: exon
? LOCATION: 7871..8102
? FEATURE:
? NAME/KEY: 3'UTR
? LOCATION: 8103..8537
? FEATURE:
? NAME/KEY: misc_RNA
? LOCATION: 8538..8878
? FEATURE:
?
? CDS
? LOCATION: join(470..583, 3258..3449, 3939..4122, 5043..5270,
? LOCATION: 5831..5944, 6633..6758, 6967..7252, 7871..8102)
US-08-206-176-3

```

	Query Match	Similarity	Score	DB 1	Length	8876;
Best Local	57;	Conservative	52.3%	Pred.	No. 18;	
Matches	57;	Conservative	4;	Mismatches	48;	Indels 0; Gaps 0;
OY	42	TTATCTTAGGGACGCTATATGGTATTTATGGCACTTTAAATWTTTGTGGGCAATTCC	101			
Db	3553	TTATTTTGTGTTTGTATTTATTTGGAAATAATTAACCAACATAACATATTTGGCCTTTGCT	3612			
OY	102	TTATCATGCTGTGACCAAGATTAAATGCTGTGCCAAATTTTGATT	150			
Db	3613	TTAGGCTTCTTCTTCTGTTCTTTCTTTGGCTTGGGCCCAAAATTTCAAAAT	3661			

Search completed: September 28, 1999, 11:34:15
Job time: 2135 sec

This Page Blank (uspto)

RESULT 2
ID M30081 standard; Protein: 81 AA.
AC M30081; 27-MAR-1998 (first entry)
DE Mouse persephin.
DE Persephin; neuturin; glial-derived neurotrophic factor; GDNF;
KW neuronal degeneration; haematopoietic cell degeneration;
KW cardiac muscle degeneration; therapy; mouse.
MS Mus musculus.
PN W09735911-AL.
PD 18-SEP-1997.
PF 14-MAR-1997; U03461.
PR 14-MAR-1996; US-615944.
PA (UNIM) UNIV WASHINGTON
PI Johnson EM, Kotzbauer PT, Lampe PA, Milderand JD;
DR WPI: 97-470818/43.
DR N-PSDB: T90802.
PT GDNF-neuturin family related growth factor, Persephin - used to
PT prevent or treat cellular, neuronal or non-neuronal, degeneration or
PT insufficiency.
PS Example 12; Page 170; 228pp: English.
CC This polypeptide sequence was deduced from a mouse persephin gene
CC (see T90802) obtained from a mouse genomic library. There is an
CC apparent anomaly in the gene sequence such that the sequence
CC encoding the RXR cleavage site and the sequence corresponding to
CC the mature persephin protein (see also W26680) are not co-linear.
CC Instead a second reading frame encodes the cleavage site and mature
CC persephin. Persephin is a novel member of the glial-derived
CC neurotrophic factor-neuturin family. Mouse and rat persephin
CC polypeptides (see W30064-68), and DNA sequences encoding them, are
CC used in claimed methods for preventing or treating neuronal
CC degeneration, haematopoietic cell degeneration and cardiac muscle
CC degeneration or insufficiency.
SQ Sequence 81 AA;

Query Match	67.48;	Score 298;	DB 26;	Length 81;
Best Local Similarity	78.28;	Pred. No. 5.50e-17;		
Matches	43;	Conservative	4;	Mismatches 8; Indels 0; Gaps 0;

```

Db      1 maagrllrlcillllslpslglwldlqeasvacklsfgkmaetrgwtvphgkns 55
        |||||  |||||  :|||||||  : || |||||  || |||||  :
QY      1 MAAGRRLRIEFLLLSLHGLGWLVDLQEA PADELSSGKMAETGRTVTPHQNNN 55

```

RESULT	3	
ID	RI2541	standard; Protein; 278 AA.
AC	RI2541;	
DT	09-SEP-1991	(first entry)
DE	Latency associated peptide (LAP) TGF-beta antagonist.	
KW	AIDS; cystic fibrosis; chronic granulation; scleroderma;	
KW	hyperproliferation; vitreoretinopathy.	
OS	Homo sapiens.	
PN	W09108291-A.	
PD	13-JUN-1991.	
PF	30-OCT-1990; U06292.	
PR	22-NOV-1989; US-441680.	
PA	(GETH) GENENTECH INC.	
PI	Levinson A, Hammonds GR, Mason A;	
DR	WPI. 91-193197/26.	
DR	N-PSDB; Q12192.	
PT	New latency associated peptide - used to treat, prevent and	
PT	diagnose conditions associated with beta TGF eg. AIDS, cystic	
PT	fibrosis, chronic granulation, etc.	
PS	Disclosure; Fig 8; 66pp; English.	
CC	Gene product may be used in the control of TGF-beta activity, where	
CC	the peptide can inactivate the TGF in circulation or tissues.	
CC	This provides a treatment for conditions brought about by an excess	
CC	of TGF-beta eg chronic granulation, hyperproliferation and fibrotic	
CC	diseases, or be used in treating TGF overproduction as a symptom of	
CC	AIDS, acute immune deficiencies, cystic fibrosis and TGF-beta	
CC	producing tumors. The LAP may also be used in site direction of	

CC disabled TGF-beta molecules, which may be reactivated at destination
SQ Sequence 278 AA;

Query Match	20.4%;	Score 90;	DB 3;	Length 278;
Best Local Similarity	43.8%;	Pred. No. 1.30e+01;		
Matches	14;	Conservative	6;	Mismatches 12;
			Indels	0;
			Gaps	0;

```

Db      6  lrllppllplwllvltgppaaaglstcktid 37
      |||:| | | :| | :| | :| | :
QY     11  llllslhlgwvlldlqefapadellssgkmae 42

```

RESULT	4
ID	R53090 standard; protein; 278 AA
NC	PF00000

DT 05-JAN-1995 (first entry)
DE Polypeptide cross-linked to hmpu-1, -2 or -3.
KW Human masking protein; Transforming growth factor beta; TGF-beta
KW inhibition; wound treatment; eczema; osteoporosis; hmpu;
KW Immune disorders; organ transplantation.
OS Homo sapiens.
PN J06092995-A.
PD 05-APR-1994.
PF 21-OCT-1991; 337508.
PR 21-OCT-1991; JP-337508.
PA (NAKA/) NAKAMURA B.
DR WPI; 94-147947/18.
PT Human masking protein and its polymer subunit - useful as
PT transforming growth factor-beta inhibitor
PT Claim 5; Page 28; 56pp; Japanese.
PC A human masking protein (hmpu) having a disulphide bond between
CC R55090 and one of the masking protein subunits hmpu-1, hmpu-2 or
CC hmpu-3 is claimed. The polypeptide can inhibit TGF-beta and is
CC useful for treatment of wounds, osteoporosis, eczema, immune
CC diseases and in organ transplantation.
SQ Sequence 278 AA;

Query Match	20.4%;	Score 90;	DB 10;	Length 278;
Best Local Similarity	43.8%;	Pred. No. 1.30e+01;		
Matches	14;	Conservative	6;	Mismatches 12;
			Indels 0;	Gaps 0

```
Db      6 lrllplllplwllvltpppaaglstcktlid 37
        | ||| | | | : | : | | | :
QY     11 lllslhlglgwldlqefapadelssgkmae 42
```

RESULT	5	
ID	R05664 standard; protein; 387 AA.	
AC	R05664;	
DT	10-MAR-1993 (revised)	
DT	14-AUG-1990 (first entry)	
DE	Simian Transforming growth factor - Betal.	
KW	HIV; AIDS; STV; vaccine; AZT; CD4; cytokines; growth factors; ds.	
KW	Key	Location/Qualifiers
FT	peptide	279..390
PN	Ep-356935-A.	
PD	7-MAR-1990.	
PF	25-AUG-1989; 115719.	
PR	25-AUG-1988; US-236698.	
PA	(ONCO-) Oncogen Ltd Partner.	
PI	Brankovan V, Lioubin M, Purchio A;	
DR	WPI; 90-068723/10.	
DR	N-PSDB; 003509.	
PT	Compns. contg. transforming growth factor beta -	
PT	used for inhibitions of HIV infection and replication in vivo.	
PS	Disclosure; Fig 1; 20pp; English.	
CC	TGF-beta may be used in vivo to prevent formation of syncytia and	
CC	inhibit HIV infection. TGF may also be used with other HIV treatments	
CC	(AZT, soluble CD4 etc.).	
SQ	Sequence 387 AA;	
Query Match	20.4%; Score 90; DB 1; Length 387;	

OM of: US-09-030-606-223 to: PIR.60.* out_format : pfs

Date: Sep 25, 1999 6:28 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:

-MODEL=frame+azp.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09030606/runat.2409199.171617.29825/app-query.fasta.1
-DB=PIR.60 -OFMT=fasta -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.500
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=bloms62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=stat -USER=US09030606 -NCPU=6 -ICPU=3 -WAIT
-THREADS=1

Search information block:

Query: US-09-030-606-223
Query length: 383
Database: PIR.60.*
Database sequences: 122810
Database length: 40065486
Search time (sec): 182.540000

seq_name	Strd Orig	EScore	len	Documentation
PIR2:S64779	68.00	140.35	0.7125	586 probable membrane protein YL028*
PIR2:S72642	66.00	140.35	1.31	301 probable ABC-type transport prd
PIR2:JC5406	64.50	141.20	2.07	181 brain and muscle Ah receptor nu
PIR2:JC5407	64.50	138.37	2.08	258 brain and muscle Ah receptor nu
PIR2:PC4288	64.50	137.98	2.09	271 brain and muscle Ah receptor nu
PIR2:B71538	64.50	134.59	2.11	410 probable hypothetical protein c
PIR2:S07744	64.50	132.04	2.13	574 NAH dehydrogenase (ubiquinone)
PIR2:JC5404	64.50	131.86	2.13	583 brain and muscle Ah receptor nu
PIR2:JC5405	64.50	131.86	2.13	626 brain and muscle Ah receptor nu
PIR2:S26965	64.50	127.39	2.16	1021 probable DNA-directed DNA poly
PIR2:S01500	64.00	134.67	2.46	352 NAH dehydrogenase (ubiquinone)
PIR2:JC0270	62.50	126.42	3.98	626 Arnt-like PAS protein/Arnt3 - m
PIR3:JC4897	61.50	121.98	5.47	805 leptin receptor. Ob-Re - rat
PIR2:JC4797	61.50	121.14	5.48	894 leptin receptor precursor - rat
PIR2:PC4184	61.50	119.05	5.52	1162 leptin receptor. Ob-Rd - rat
PIR2:S44840	61.00	135.70	6.11	124 K02D10.2 protein - Caenorhabdit
PIR1:WMTMGM	61.00	115.23	6.50	1608 183K protein - tobacco mild gr
PIR2:D64108	60.00	122.32	8.64	488 glucanate transport protein hom
PIR2:T02568	60.00	116.68	8.79	989 hypothetical protein r16824.2 -
PIR2:C71637	59.50	124.88	9.98	304 rod shape-determining protein m
PIR2:S57058	59.50	114.46	10.30	1121 probable membrane protein YL02
PIR2:PC4432	59.00	119.12	11.83	537 hypothetical protein M1060 - M
PIR2:D64108	58.50	121.39	13.68	347 hypothetical protein R1338 - R1
PIR2:B71690	58.50	121.16	13.69	357 lipopolysaccharide 1,2-n- acety
PIR2:A65163	58.50	119.51	13.76	439 preprotein translocase secy - S
PIR2:S27155	58.50	113.62	14.01	918 vacuolar membrane protein PEP3
PIR2:A41943	58.00	124.11	15.81	212 phosphoglycerate mutase 1 - Aqu
PIR2:A70471	58.00	123.27	15.85	235 probable glucose-6-phosphate 1
PIR2:EV0127	57.50	102.56	16.88	3149 BHLF1 protein - human hepesv1
PIR1:Q00B8	57.50	122.48	18.50	223 urec1 DNA glycosylase (ung) hc
PIR2:E70106	57.50	118.12	18.75	385 Orec1 - Bacillus firmus
PIR2:JC39826	57.50	114.09	18.98	638 somatotropin receptor precursor
PIR2:A33991	57.50	114.09	18.98	638 somatotropin receptor, hepatic
PIR2:S04530	57.50	113.99	18.99	646 ABC transporter (permease) homc
PIR2:B70001	57.50	110.44	19.19	1008 acriflavin resistance protein
PIR2:E71177	57.50	106.36	19.43	1680 complement C5 precursor - mous
PIR1:C5MS	57.00	119.95	21.72	263 spermidine/putrescine ABC trans
PIR2:G70179	57.00	125.41	24.88	114 hypothetical protein PH0953 - F
PIR2:D71086	56.50	120.27	25.27	217 hypothetical protein PH1936 - F
PIR2:G64642	56.50	118.40	25.42	274 conserved hypothetical integrat
PIR2:A45033	56.50	110.63	26.02	725 myelin transcription factor 1 -
PIR2:A71704	56.50	110.01	26.07	784 ATP-dependent proteinase LA (ld

PIR1:C5HU + 56.50 103.94 26.56 1676 | complement C5 precursor - h
PIR2:E71495 + 56.00 120.19 29.44 188 | probable peptidoglycan-assoc
PIR2:B70693 + 56.00 117.95 29.65 249 | probable echA16 protein - My

seq_name: PIR2:S64779

seq_documentation_block:

Probable membrane protein YL028* - Yeast (Saccharomyces cerevisiae)
N: Alternate names: hypothetical protein L0939
C: Species: Saccharomyces cerevisiae
C: Date: 01-Aug-1995 #SequenceRevision 24-May-1996 #text_change 14-Nov-1997
C: Accession: S64779
R: Duesetoeft, A.; Floeth, M.; Heuss-Neitzel, D.; Hilbert, H.; Moestl, D.
Submitted to the Protein Sequence Database, May 1996
A: Reference number: S64775
A: Accession: S64779
A: Molecule type: DNA
A: Residues: 1-586 <DUE>
A: Cross-references: EMBL:Z73133; NID:q1360209; PID:e245461; PID:q1360210; MIPS:YL028
A: Experimental source: strain S288C
C: Genetics:
A: Map position: 12L
C: Keywords: transmembrane protein
F: 146-162/Domain: transmembrane #status predicted <TM1>
F: 179-195/Domain: transmembrane #status predicted <TM2>
F: 214-230/Domain: transmembrane #status predicted <TM3>
F: 240-256/Domain: transmembrane #status predicted <TM4>
F: 273-289/Domain: transmembrane #status predicted <TM5>
F: 306-322/Domain: transmembrane #status predicted <TM6>
F: 369-385/Domain: transmembrane #status predicted <TM7>
F: 389-405/Domain: transmembrane #status predicted <TM8>
F: 416-432/Domain: transmembrane #status predicted <TM9>
F: 457-473/Domain: transmembrane #status predicted <TM10>
F: 489-505/Domain: transmembrane #status predicted <TM11>
F: 553-569/Domain: transmembrane #status predicted <TM12>

alignment_scores:

Quality: 68.00 Length: 92
Ratio: 1.361 Gaps: 5
Percent Similarity: 46.739 Percent Identity: 26.087

alignment_block:

US-09-030-606-223 x S64779 ..

Align seg 1/1 to: S64779 from: 1 to: 586

```

91 GGGGCAATTCCTTACATGCTTGACAAAGATTAAATGCTGCCAAA 140
||| .....|||.....|||
414 GYGLUUEUPROTYRILEALALELLELLEGLYMETHEVALCYSAIAI 430
||||.....|||.....|||
141 TTTTGATATTTTATTTGAGACTCTTATCAAAAGTAAGTCCCAAGA 190
||||.....|||.....|||
430 APhellleTTPlyrmetaspnsapTyrleuLysArgCys.....Arga 445
||||.....|||.....|||
191 AGCTTAGCAATTAGTA..... 207
||||.....|||.....|||
445 lalysGLYlSleValPProGLuAlaArgleuTyrAlaMetValIleAla 461
||||.....|||.....|||
208 .....GTGTCCCM.....TCACGTGTTGGAGTGTCTATTCAAGAT 248
||||.....|||.....|||
462 gLYthralPheProIleGLYleleuTTP..... 471
||||.....|||.....|||
249 TTTGATTCCTGATGACATTAAT.....ATTGAA 280
||||.....|||.....|||
472 .....PhecysTTPheGLYtyrTyrProHISLysIleHIStrMetValP 487
||||.....|||.....|||
281 CTTTGCGGGGAANAGTATAGA 306
||||.....|||.....|||
487 rothrValGLYglYAlaPheIleGLY 495
||||.....|||.....|||
seq_name: PIR2:S72642

```



```

alignment_scores:
    Quality: 64.00      Length: 142
    Ratio: 0.889      Gaps: 8
    Percent Similarity: 50.704      Percent Identity: 22.535

alignment_block:
US-09-030-606-223 x S01500 ..

Align seg 1/1 to: S01500 from: 1 to: 352

21 ACAATTCCTCATTCAGAAAAATATCTTAGGACGTGATATGGTAATAT 70
||||| ::: ::||| ::|||::|||::|||::|||
232 ThrIleSerGlnIleuSerProIleSerValAlaIleuValIleuValIle 248
71 GGTCAATTAATTAATTTTTRTTRTKTGCGGCATTTCTT 102
|||||::||| |||::|||
248 ILeuSerIleu.....GlyGlyIleu.ProProIleuThrGlyIlePheIle 261

```

262 LeuysPmethrSerLeutyPhel euValaAlaansnphellele 278
114 GACAAGATTAATAATCTG.....TGC AAAATTTGTAATTATTGG 157
178 uSerlelleketlelleGlyasleuGlnAspTyPhepheyrleuA 295
158 AGACTTCTATCAAAAGTATGCTGCCAAAGGAAGTCTAAGAAATTAGTA 207

[illegible]

```

Ant-like PAS protein,Ant3 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_changed
C:Accession: JE0270
C:Taxahata, S.; Sogawa, K.; Kobayashi, A.; Ena, M.; Mimura, J.;
Biophys. Res. Commun. 248, 789-794, 1998
A:Title: Transcriptionally active heterodimer formation of an Ar
A:Reference number: JE0270
A:Accession: JE0270
A:Molecule type: mRNA
A:Residues: 1-626 <TAK>
A:Cross-references: DDBJ:AB014494

Alignment scores:
      Quality: 62.50      Length: 76
      Ratio: 1.689      Gaps: 1
Percent Similarity: 48.684      Percent Identity: 23.684

Alignment_block:
US-09-030-606-223/rev x JE0270
..

```

```

292 CCCCCCAAGTTAAATATATGTCATTCAGGAAATCAAACTTT 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
57 ProHISGLYArgLeuGIuYrAlaGLuHISGLInGLYArgILEuSAsnAL 73
242 TAGAATAGCAGACTCCAAACAACTGAKGGACACTACTATTCCTTACA 193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
73 aArgYLuaLHISserGLn..... 79
192 CTTCCTTTGGCAGCATTACTTTGTATAGAGAGTCTCCAAATAAATACAA 143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
80 ..... 89
142 AATTTTGGCAGACACATTTTAACTTGTCTCAAGACAATGTAGAGAAATGCC 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90 SerPheILeAspGLueuAspLysLeuThrValProThrcysAsnAlaMetse 106
92 CCMAAAATATTAATTAATTCACCATAT 65
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
106 rArgYLysLeuAspLysLeuThrValLeu 115
seq_name: p1r3:JC4897
seq_documentation_block:
leptin receptor, OD-Re - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 03-Sep-1998 #sequence_revision 03-Sep-1998 #text_change 03-Sep-1998
C:Accession: J04897
R:Takaya, K.; Ogawa, Y.; Isse, N.; Okazaki, T.; Satoh, N.; Masuzaki, H.; Mori, K.; Tamur
Biochem. Biophys. Res. Commun. 225, 75-83, 1996
A:Title: Molecular cloning of rat leptin receptor isoform complementary DNAs-identificat
A:Reference number: J04897; MUID:96332408
A:Accession: J04897
A:Status: preliminary
A:Residues: 1-805 <TAK>
A:Cross-references: DDBJ:D85559

```

```

alignment_scores:
Quality: 61.50 Length: 111
Ratio: 0.992 Gaps: 8
Percent Similarity: 55.856 Percent Identity: 31.532

```

```

alignment_block:
US-09-030-606-223 x JC4897 ..

```

```

Align seg 1/1 to: JC4897 from: 1 to: 805

```

```

17 AAAACAATCT.....TCATTCAGAAAATATCTTA..... 49
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
667 LysAsnAspSerLeuGlySerValArgArgTYrValValLysHISArgTh 683
50 .....GGGACT.....GATATTGGTAAATATAGTCATTTA 80
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
683 rAlaHISAsnGLYThrTrpSerGLnAspValGLYAsn.GlnThrAsnLeu 699
81 ATWRPTTGTGGGCGATTCCTTACAT.....TGCTTGA 115
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
700 ThrPheLeuTrpAlaGLuSerAlaHISThrValThrValLeuAlaILEAs 716
116 CAAGATTAATAATCTGTGCGCAAAATTTGTATTTATTTGGAGACTTCT 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
716 nSerILleGLYAlaSerLeuValAsnPhenAsnLeuThrPheSerTrpProM 733
166 TATCAAAAGTAATGCTGCCAAGAGCTCAAGGAATTAAGTGTCC 215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
733 eSerLYsVal.....AsnAlaValGLnSerLeuSerAlaTYrPro 746
216 MTCAGTTGTTGGAGTGTCTATCTTAAGATTTTGAATTTCTCGAAG 265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
747 LeuSerSer...SerCYsValILEuSerTrpThrLeuSerProAsnA 762
266 ACAATTATATTTAACTTTGTTGGGGGAA 294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

762 sPTYrSerLeuLeuTYrLeuValILEGLu 771

```

```

seq_name: p1r2:JC4797

```

```

seq_documentation_block:
leptin receptor precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 15-Oct-1995 #sequence_revision 16-Aug-1996 #text_change 17-Mar-1999
C:Accession: J04797; J04896
R:Ida, M.; Murakami, T.; Ishida, K.; Mizuno, A.; Kuwajima, M.; Shima, K.
Biochem. Biophys. Res. Commun. 222, 19-26, 1996
A:Title: Phenotype-linked amino acid alteration in leptin receptor cDNA from Zucker f
A:Reference number: J04797; MUID:96212906
A:Accession: J04797
A:Molecule type: mRNA
A:Residues: 1-894 <IID>
A:Cross-references: DDBJ:D84125; NID:91374707; PID:d1012905; PID:g1374708
A:Experimental source: adipose cell
R:Takaya, K.; Ogawa, Y.; Isse, N.; Okazaki, T.; Satoh, N.; Masuzaki, H.; Mori, K.; Ta
Biochem. Biophys. Res. Commun. 225, 75-83, 1996
A:Title: Molecular cloning of rat leptin receptor isoform complementary DNAs-identifi
A:Reference number: J04895; MUID:96332408
A:Accession: J04896
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-894 <TAK>
A:Cross-references: DDBJ:D85557
C:Genetics: Fa
C:Keywords: appetite; transmembrane protein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-894/Product: leptin receptor #status predicted <MAT>
F:840-860/Domain: transmembrane #status predicted <TM>

```

```

alignment_scores:
Quality: 61.50 Length: 111
Ratio: 0.992 Gaps: 8
Percent Similarity: 55.856 Percent Identity: 31.532

```

```

alignment_block:
US-09-030-606-223 x JC4797 ..

```

```

Align seg 1/1 to: JC4797 from: 1 to: 894

```

```

17 AAAACAATCT.....TCATTCAGAAAATATCTTA..... 49
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
667 LysAsnAspSerLeuGlySerValArgArgTYrValValLysHISArgTh 683
50 .....GGGACT.....GATATTGGTAAATATAGTCATTTA 80
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
683 rAlaHISAsnGLYThrTrpSerGLnAspValGLYAsn.GlnThrAsnLeu 699
81 ATWRPTTGTGGGCGATTCCTTACAT.....TGCTTGA 115
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
700 ThrPheLeuTrpAlaGLuSerAlaHISThrValThrValLeuAlaILEAs 716
116 CAAGATTAATAATCTGTGCGCAAAATTTGTATTTATTTGGAGACTTCT 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
716 nSerILleGLYAlaSerLeuValAsnPhenAsnLeuThrPheSerTrpProM 733
166 TATCAAAAGTAATGCTGCCAAGAGCTCAAGGAATTAAGTGTCC 215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
733 eSerLYsVal.....AsnAlaValGLnSerLeuSerAlaTYrPro 746
216 MTCAGTTGTTGGAGTGTCTATCTTAAGATTTTGAATTTCTCGAAG 265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
747 LeuSerSer...SerCYsValILEuSerTrpThrLeuSerProAsnA 762
266 ACAATTATATTTAACTTTGTTGGGGGAA 294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
762 sPTYrSerLeuLeuTYrLeuValILEGLu 771

```


seq_name: plr2:PC4184

seq_documentation_block:

leptin receptor, Ob-Rb - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 15-Aug-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-1999
 C:Accession: J04895; PC4184
 R:Takaya, K.; Ogawa, Y.; Isse, N.; Okazaki, T.; Satoh, N.; Masuzaki, H.; Mori, K.; Tamura, K.; Koyama, T.; Ishida, K.; Mizuno, A.; Kuwajima, M.; Shima, K.
 Biochem. Biophys. Res. Commun. 225, 75-83, 1996
 A:Title: Molecular cloning of rat leptin receptor isoform complementary DNAs-identification
 A:Reference number: J04895; M01D:96332408
 A:Accession: J04895
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1162 <TAK>
 A:Cross-references: DDBJ:D85586; NID:91526441; PID:d1013515; PID:91526442
 R:Ida, M.; Murakami, T.; Ishida, K.; Mizuno, A.; Kuwajima, M.; Shima, K.
 Biochem. Biophys. Res. Commun. 224, 597-604, 1996
 A:Title: Substitution at codon 269 (glutamine-proline) of the leptin receptor (OB-R) cDN
 A:Reference number: PC4184; M01D:96295531
 A:Accession: PC4184
 A:Molecule type: mRNA
 A:Residues: 840-1162 <IID>
 A:Cross-references: DDBJ:D84550
 C:Comment: This receptor is obese-phenotype-linked mutant, found in the Zucker (fa/fa) m
 C:Keywords: appetite; transmembrane protein
 F:840-860/Domain: transmembrane status predicted <TM>
 F:861-1162/Domain: Intracellular status predicted <INT>

alignment_scores:

Quality:	61.50	Length:	111
Ratio:	0.992	Gaps:	8
Percent Similarity:	55.856	Percent Identity:	31.532

alignment_block:

US-09-030-606-223 x PC4184

Align seg 1/1 to: PC4184 from: 1 to: 1162

```

17  AAAAACAATTCT.....TCATCGAAAAAATATCTTA..... 49
|||||:|||||  |||  |||:|  |||:|  |||:|
667  LysAsnAspSerLeuCySerValAlaGlyTyrValValLysHisArgTh 683
50  .....GGGACT.....GATATGGTAATATGTCATTTA 80
|||||  |||  |||:|  |||:|  |||:|  |||:|
683  ValHisAsnGlyThrTrpSerGlnAspValGlyAsn.GlnThrAsnLeu 699
81  ATWRTTTRTKTGCGCATTTCTTACAT.....TGTCCTGA 115
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
700  ThrPheLeuTrpAlaGlnSerAlaHisThrValThrValLeuAlaIleAs 716
116  CAAGATTAAATGTCGTGTCGCAAAATTTGATTTTATTTGAGACTTCT 165
|||||  |||  |||:|  |||:|  |||:|  |||:|
716  nSerIleGlyIleSerLeuValAsnPheAsnLeuThrPheSerTrpPro 733
166  TATCAAAAGTATGCTGCCAAGAGCTCTAAGCATTAAGTGTCC 215
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
733  etSerLysVal.....AsnAlaValGlnSerLeuSerAlaIleAsnPro 746
216  MTGACTGTTGAGAGTGTGCTATTTCTAAAGATTGATTTCTCGAATG 265
|||||  |||  |||:|  |||:|  |||:|  |||:|
747  LeuSerSer...SerCysValIleLeu.SerTrpThrLeuSerProAsn 762
266  ACAATTATATTTTAACCTTGGTGGGGGAA 294
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
762  spTySerLeuLeuTyrLeuValIleGln 771

```

This Page Blank (uspto)

OM of: US-09-030-606-223 to: SwissProt_37.* out_format: pfs
Date: Sep 25, 1999 4:46 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:

-MODEL=framedb -n2p -model -DEV=xlp
-O=/gen2.1/USPROT/US09030606/runat_24091999_171618_29883/app-query.fasta.1
-DB=SwissProt_37 -OPMT=fastan -SUFFIX=esp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsim62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=score
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606
-NCPD=6 -ICPD=3 -WAIT -THREADS=1

Search information block:

Query: US-09-030-606-223
Query length: 383
Database: SwissProt_37.*
Database sequences: 77977
Database length: 28268293
Search time (sec): 155.430000

Sequence: Stnd Orig Zscore EScore Len
SwissProt_37:NU5M_PARTE + 64.50 129.50 2.08
SwissProt_37:BMAL_HUMAN - 64.50 129.50 2.08
SwissProt_37:DPOK_NEUCR + 64.50 124.76 2.13
SwissProt_37:NU2M_STRPU + 64.00 132.44 2.37
SwissProt_37:YMO2_CAEEL - 61.00 133.66 5.60
SwissProt_37:RRPO_TYGMV - 61.00 112.82 6.26
SwissProt_37:YJ09_YEAST - 59.50 112.22 9.69
SwissProt_37:REAK_ECOLI - 58.50 119.17 12.48
SwissProt_37:SECK_SNMP7 + 58.50 117.49 13.59
SwissProt_37:PEP3_YEAST + 58.50 111.49 13.00
SwissProt_37:CEB2_HELAM + 58.00 115.18 14.74
SwissProt_37:TEGU_EBV + 58.00 100.29 15.94
SwissProt_37:UNG_BORBU - 57.50 120.64 16.55
SwissProt_37:YCT2_BACFI + 57.50 116.20 16.94
SwissProt_37:GHR_HUMAN + 57.50 112.09 17.31
SwissProt_37:CO5_MOUSE + 57.50 104.22 18.05
SwissProt_37:YCI0_KLEPN + 57.00 114.14 19.80
SwissProt_37:MYT1_HUMAN + 56.50 108.69 23.55
SwissProt_37:CO5_HUMAN + 56.50 101.88 24.41
SwissProt_37:EAB1_CHICK + 56.00 121.81 25.40
SwissProt_37:PSD5_HUMAN + 56.00 110.47 26.96
SwissProt_37:IMH3_CANAL - 56.00 110.20 27.00
SwissProt_37:RPOA_SPTOL + 55.50 112.62 30.82
SwissProt_37:SPAS_SALT + 55.50 112.12 30.90
SwissProt_37:YC24_CYAPA + 55.50 109.59 31.31
SwissProt_37:EXO1_YEAST + 55.50 106.60 31.81
SwissProt_37:CNB3_HUMAN - 55.50 102.86 32.45
SwissProt_37:RPA1_MOUSE + 55.00 99.33 33.06
SwissProt_37:YPI1_CAEEL + 55.00 115.80 33.03
SwissProt_37:YRBE_HAEIN + 55.00 113.47 35.46
SwissProt_37:CYST_SINY3 + 55.00 112.72 35.60
SwissProt_37:US20_HCMVA + 55.00 111.27 35.87
SwissProt_37:US20_HCMVT + 55.00 111.27 35.87
SwissProt_37:INOB_PSEAE + 55.00 106.34 36.78
SwissProt_37:MCM4_MOUSE + 55.00 103.75 37.33
SwissProt_37:COBL_METJA + 54.50 114.02 40.87
SwissProt_37:Y184_MYCPN + 54.50 110.66 41.60
SwissProt_37:QUEA_AOUAN + 54.50 110.28 41.68
SwissProt_37:RFBX_ECOLI + 54.50 108.52 42.07
SwissProt_37:FUMH_YEAST + 54.50 107.60 42.37
SwissProt_37:ACM1_BOVIN + 54.50 105.66 42.71
SwissProt_37:G6PC_TOBAC - 54.50 105.62 42.72

SwissProt_37:UVRC_HAEIN - 54.50 105.40 42.77 609 | P44489 haemophilus influe
SwissProt_37:POL2_TBARS + 54.50 98.89 44.27 1357 | P14547 tomato black ring
SwissProt_37:TD2_YEAST - 54.50 98.59 44.34 1407 | P23255 saccharomyces cer

seq.name: SwissProt_37:NU5M_PARTE

seq.documentation_block:

ID NU5M_PARTE STANDARD: PRT: 570 AA.
AC P15584:
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).
GN ND5 OR NDH5.
OS PARAMECIUM TETRAURELLA.
OC MITOCHONDRION.
OC EUKARYOTA; ALVEOLATA; CILIOPHORA; NASSOPHOREA; PENICULIDA; PARAMECIUM.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-STOCK 51.
RX MEDLINE: 90174913.
RA PRITCHARD A.E., SEITZNER J.J., MAHALINGAM R., SABLE C.L.,
RA VENTRI S.E., COMINGS D.J.;
RT "Nucleotide sequence of the mitochondrial genome of Paramecium";
RL NUCLEIC ACIDS RES. 18:173-180(1990).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@sdb.ch).
CC -----
DR EMBL: X15917; G13276; -
DR PIR: S07744; S07744.
DR PFAM: PF00361; oxidored_g1: 1.
DR PFAM: PF00662; oxidored_g1_N: 1.
KW OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION.
SQ SEQUENCE 570 AA: 65196 MW: 8580309F CRC32;

alignment_scores:

Quality: 64.50 Length: 119
Ratio: 1.093 Gaps: 7
Percent Similarity: 49.580 Percent Identity: 25.210

alignment_block:

US-09-030-606-223 x NU5M_PARTE ..

Align seg 1/1 to: NU5M_PARTE from: 1 to: 570

105 CATGTGTCGACAGATTAAATG...TCGTGCGCAAA... 140
||||||..... |||.....
203 HtsleuylserglumetlyslenglySerThrProInlileasnserTr 219
.....TTTGTATTATTATTTGGAGACTTGTATCAAA... 173
141
219 psmleuilleserhethesleuileuphealialaphelvallysserlaag 236
|||.....
174GTATGCTGCGCAAGAGAGCTAAGAAATTAGAGT 209
236 lnpheglyphenhlsvaltrpneuProaspsermet.....GluAlapro 250
210 GTTCCGCTTCACTTGTGG... 229
|||.....
251 ValProhlaaeralaleuilehsserAlatThrleuValserlaaglyva 267
230 GTGTGCTATTCTAAAGATTGATTTCCTGGAATGACAAATATATTATTA 279
|||.....
267 lpheluleilemetatrpheyltrProileleuGluLeuSerleutyrphel 284

280 AC.....TTGGTGGGGAAN 296
 284 yslseuValThrAlaLeuValGlyAlaLeuThrAlaLeuAlaGlyGlyLeu 300
 297 AGTTATAGACACAGCTCTTCTGATGACTGTAATTAATCTTTA 346
 301 Ser.....AlaValPheGlnThrAspLeuLysLysIleLeuAlaTy 314
 347 TTGCAC 353
 314 rserThr 316

seq_name: SwissProt_37:BMAL_HUMAN

seq_documentation_block:

ID BMAL_HUMAN STANDARD. PRT: 583 AA
 AC 000327; 000313; 000314; 000315; 000316; 000317; 099631; 099649;
 DT 15-DEC-1998 (REL. 37, CREATED)
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE BMAL1 PROTEIN (MEMBER OF PAS PROTEIN 3) (MOP3) (BHLH-PAS PROTEIN JAP3).
 GN BMAL1 OR ARNTL.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS BMAL1 A-F).
 RC TISSUE-BRAIN;
 RX MEDLINE: 97289529.
 RA IKEDA M., NOMURA M.;
 RT "DNA cloning and tissue-specific expression of a novel basic helix-loop-helix/PAS protein (BMAL1) and identification of alternatively spliced variants with alternative translation initiation site usage."
 RT BIOCHEM. BIOPHYS. RES. COMMUN. 233:258-264(1997).
 RL [2]
 RN RP SEQUENCE FROM N.A. (ISOFORM MOP3).
 RC TISSUE-FETAL BRAIN;
 RX MEDLINE: 97236817.
 RA HOGNESECH J.B., CHAN W.K., JACKIM V.H., BROWN R.C., GU Y.-Z.,
 RA PRAT-GRANT M., PERDEW G.H., BRADFIELD C.A.;
 RT "Characterization of a subset of the basic-helix-loop-helix-PAS superfamily that interacts with components of the diroxin signaling pathway."
 RT J. BIOL. CHEM. 272:8581-8593(1997).
 RL [3]
 RN RP SEQUENCE FROM N.A. (ISOFORM BMAL1B).
 RA TIAN H., RUSSELL D.W., MCKNIGHT S.L.;
 RA SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1 SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH AN OTHER BHLH PROTEIN. INTERACTS WITH HSP90; WITH AIR IN VITRO, BUT NOT IN VITRO.
 CC -1 SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1 TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE ADULT BRAIN, SKELETAL MUSCLE AND HEART.
 CC -1 ALTERNATIVE PRODUCTS: AT LEAST SEVEN ISOFORMS: BMAL1A (SHOWN HERE), BMAL1B/JAP3, BMAL1C, BMAL1D, BMAL1E, BMAL1F AND MOP3; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1 SIMILARITY: CONTAINS A PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
 CC -1 SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF TRANSCRIPTION FACTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation-CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D89722; E316533; -
 CC EMBL: AB000812; D1020725; -

DR EMBL: AB000813; D1020726; -
 DR EMBL: AB000814; D1020727; -
 DR EMBL: AB000815; D1020728; -
 DR EMBL: AB000816; D1020729; -
 DR EMBL: U51627; G1695803; -
 DR EMBL: U60415; G1698576; -
 DR MIM: 602550; -
 DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
 DR PFM: PF00010; BHLH; 1.
 DR PFM: PF00983; PAS; 1.
 KW REPEAT; DNA-BINDING; NUCLEAR PROTEIN; TRANSCRIPTION REGULATION;
 KW ALTERNATIVE SPLICING.
 FT DNA_BIND 30 42
 FT DOMAIN 43 83
 FT REPEAT 103 170
 FT REPEAT 285 351
 FT DOMAIN 358 401
 FT VARSPLIC 1 4
 FT VARSPLIC 1 16
 FT VARSPLIC 181 181
 FT VARSPLIC 182 583
 FT VARSPLIC 231 348
 FT VARSPLIC 235 258
 FT VARSPLIC 259 583
 FT VARSPLIC 400 483
 FT VARSPLIC 484 583
 FT CONFLICT 26 26
 FT CONFLICT 80 80
 FT CONFLICT 130 130
 FT CONFLICT 216 216
 FT CONFLICT 221 221
 FT CONFLICT 375 375
 FT CONFLICT 470 471
 SQ SEQUENCE 583 AA; 64206 MW; 37642A96 CRC32;
 alignment_scores:
 Quality: 64.50 Length: 76
 Ratio: 1.697 Gaps: 1
 Percent Similarity: 50.000 Percent Identity: 23.684
 alignment_block:
 US-09-030-606-223/rev x BMAL_HUMAN ..
 Align seg 1/1 to: BMAL_HUMAN from: 1 to: 583
 292 CCCCACCAAGTTAAATATATATCTGATCCAGGAATCAAAATCTT 243
 14 ProHisGlyArgIleuGlnIuYrThGlnHisGlnGlyArgIleLysSmaI 30
 242 TAGAATAGACACTCCAAACAGTAGGAGAACACTACTAATCTCTTAA 193
 30 aArgGlnAlaHisSerGln..... 36
 192 CTTCCTTTGGCAGCATTTGATTAAGATCTCCAAATAATAATACA 143
 37 46
 142 AATTGGCAGACATTTAATCTGTCAAGCATGTAGAGAAATGCC 93
 47 SerPheIleAspIuLeuAlaSerLeuValProThrCysAlaMetse 63
 92 CCMAAYAYWATTAAATGACCAATAAT 65
 63 rArgLysLeuAspLysLeuThrValLeu 72

seq_name: SwissProt_37:DPOK_NEUCR

seq_documentation_block:

ID DPOK_NEUCR STANDARD; PRT; 1021 AA.
AC P33537;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DE 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PROBABLE DNA POLYMERASE (EC 2.7.7.7).
OS NEUROSPORA CRASSA.
OG MITOCHONDRION.
OC PLASMID MARANHAR.
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PYRENOMYCETES;
OC SORDARIALES; SORDARIACEAE; NEUROSPORA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AAREY-1E;
RX MEDLINE; 93046810.
RA COURT D.A., BERTRAND H.;
RT "Genetic organization and structural features of maranhar, a senescence-inducing linear mitochondrial plasmid of Neurospora crassa."
RT CURR. GENET. 22:385-397(1992).
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
CC N PYROPHOSPHATE + DNA(N).
CC -1- THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. STRONG, TO
CC DNA POLYMERASE OF OTHER FUNGAL AND PLANTS MITOCHONDRIAL PLASMIDS.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X55361; E246763; -.
DR PIR; S26985; S26985.
DR PROSITE; PS00116; DNA-POLYMERASE_B; 1.
KM TRANSFERASE; DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION;
KM DNA-BINDING; PLASMID; MITOCHONDRION.
SQ SEQUENCE 1021 AA; 119075 MW; 2ED87AC0 CRC32;

alignment_scores:

Quality: 64.50 Length: 72
Ratio: 1.613 Gaps: 3
Percent Similarity: 55.556 Percent Identity: 27.778

alignment_block:

US-09-030-606-223 x DPOK_NEUCR ..

Align seg 1/1 to: DPOK_NEUCR from: 1 to: 1021

138 AAATTGTTGATTTATTTGAGACTCTTATCAAAAGTAATGCTGCCAAA 187
|||||
383 LysPheAspLeu.....AsnHeiLeileIysIleLeuValGlnGI 396
188 GGAGTCTAAGGAATTAGTAGTCTCCMTCACTGTTGGAGAGTGTGCTA 237
|||
396 upheval.....ValGluIysIleI 403
238 TTTCAAAAGATTTGATTTCTCGAATGACAATATATTTTAACTTT... 284
|||||
403 IeserLysAspLeuAspIleLeuSerIleIysIleSerTyrIlyshegu 419
285GGTGGGGAANAAGTTATATAGACACACAGCTTCACTTCTGA 325
|||||
420 ProIysLysLysGlyGlyLysAlaGlnIurghIsthrIleThrIleAlaAs 436

326 TACTTGTAATTAATC 341
|:::|:::|:::|:::
436 pSerCysArgLeuLeu 441

seq_name: SwissProt_37:NU2M_STRPU

seq_documentation_block:

ID NU2M_STRPU STANDARD; PRT; 352 AA.
AC P15549;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 2 (EC 1.6.5.3).
OS ND2.
OS STRONGYLOCENTROTUS PURPURATUS (PURPLE SEA URCHIN).
OG MITOCHONDRION.
OC EUKARYOTA; METAZOA; ECHINODERMATA; ECHINOZOA; ECHINOIDEA;
OC EUECHINOIDEA; ECHINACEA; ECHINOIDA; STRONGYLOCENTROTIDAE;
OC STRONGYLOCENTROTUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89011951.
RA JACOBS H.T., ELLIOTT D.J., MATH V.B., FAROUHARSON A.;
RT "Nucleotide sequence and gene organization of sea urchin mitochondrial DNA."
RT J. MOL. BIOL. 202:185-217(1988).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X12631; G13666; -.
DR PIR; S01500; S01500.
DR PFM; PF00361; oxidored_q1; 1.
KM OXIDOREDUCTASE; NAD; UBIQUINONE; MITOCHONDRION; TRANSMEMBRANE.
SQ SEQUENCE 352 AA; 38653 MW; F3D3D39D CRC32;

alignment_scores:

Quality: 64.00 Length: 142
Ratio: 0.889 Gaps: 8
Percent Similarity: 50.704 Percent Identity: 22.535

alignment_block:

US-09-030-606-223 x NU2M_STRPU ..

Align seg 1/1 to: NU2M_STRPU from: 1 to: 352

21 ACATTTCTCATTCAGAAAATTTATCTAGGACTGATATGTAATTAT 70
|||||
232 ThrIleSerGlnLeuSerProIleSerValAlaLeuValLeuLeuValMe 248
71 GGCAATTTAATWTRRTTKGGGCACTTCTCT 102
|:::|:::|
248 tleuSerLeu.....GlyGlyLeu.ProProLeuThnglyPheIle 261
103TACATGTGCTT 113
262 LeuIysPheThrSerLeuTyrPheLeuValAlaAsnAsnPhelIleIle 278
114 GACAAGATTAATAATGCTG.....TGCCAAATTTGATTTATTTGCG 157
|:::|:::|
278 IuSerIleIleIleIleIleGlyAsnLeuGlnAspTyrPhePheTyrIleuA 295
158 AGACTTCTTATCAAAAGTAATGCTGCCAAAGAGAGTCAAGAAATTAAGTA 207
::|:::|:::|:::|:::|

Ratio: 1.605 Gaps: 3
Percent Similarity: 53.521 Percent Identity: 25.352

alignment_block:

US-09-030-606-223/rev x RPO_TMGMV ..

Align seg 1/1 to: RPO_TMGMV from: 1 to: 1608

```

238 ATACACACCTCCAAAGAGTGAAGGAGACTACTAATTCCTTAGACTTC 189
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1320 ValTyrHisSerLysGlnIleAsnGlyIleLeu.....AlaG1 1332
188 CTTGGCAGCATTA..... 173
1332 yPhseSerGluLeuThrArgLeuLeuGluAlaPheAspSerLysLysP 1349
172 .....TTTGATPAGAGAGTCTCCAAATAAATACAAATTTGGCACA 131
1349 heLeuPhePheThrArgLysThrProGluGlnIleGlnGlnPheSer 1365
130 GACATTTTAACTTGTCAAGCAATGTAGGAATGCCCAAAAYAT 81
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1366 AspLeu.....AspSerHisValPrometaspValle 1376
80 TAAATTGACATA 68
    :|||:|||||:
1376 uGluLeuaspIle 1380

```

seq_name: SwissProt_37:YJ09_YEAST

seq_documentation_block:

```

ID YJ09_YEAST STANDARD; PRT; 1121 AA.
AC P47107;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOHENICAL 127.4 KD PROTEIN IN RAD26-GEPI INTERGENIC REGION.
GN YJR039M OR J1614.
OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCETES; SACCAROMYCETALES;
OC SACCAROMYCETACEAE; SACCAROMYCETES.
RN [1]
RA HUANG M.-E.; CHUAT J.-C.; GALIBERT F.;
RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: 249538; G1015689; -
CC DR HYPOHENICAL PROTEIN.
CC KW
SQ SEQUENCE 1121 AA; 127438 MW; F5B4818A CRC32;

```

alignment_scores:

Quality: 59.50 Length: 79
Ratio: 1.352 Gaps: 3
Percent Similarity: 55.696 Percent Identity: 27.848

alignment_block:

US-09-030-606-223/rev x YJ09_YEAST ..

Align seg 1/1 to: YJ09_YEAST from: 1 to: 1121

```

283 AAGTTAAATTAATTCATTCACGAAATCAAAATCTTTAGAAATAC 234
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
885 GlnLeuAsnTyrSerValHisAspAsnLysPheSerIleGluGlnValSe 901

```

```

233 ACATCCAAACAGTGAAGGAGACTACTAATTCCTTAGACTTCCTTG 184
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
901 AsnArgLeuAsnIleSerGlyIle.....ThilerHrs 913
183 GCACATTAATTCCTTAGTAAAG.....AGTCTCCAAATAAATACAA 143
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
913 erSerIlePhePheAspLysArgLysAlaLysMetAlaArgLysGlnGln 929
142 AATTTGGACA.....GACATTTAATCTTGTCACAGACA 108
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
930 AsnIleGlyThrPheValTyrLeuGluGluMetIleLeuLeuaspValAr 946
107 ATGTAAGAAATGCCCAAAAYATTAATTAATGACC 71
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
946 glyGlyValMetArgPheAsnValIleHisThrThr 958

```

seq_name: SwissProt_37:RFAK_ECOLI

seq_documentation_block:

```

ID RFAK_ECOLI STANDARD; PRT; 357 AA.
AC P27242;
DT 01-ANG-1992 (REL. 23, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE LIPOPOLYSACCHARIDE 1,2-N-ACETYLGLUCOSAMINETRANSFERASE (EC 2.4.1.56).
GN RFAK.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RN [1]
RA SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE: 92325067.
RA KLENA J.D., PRADEL E., SCHNAITMAN C.A.;
RT "Comparison of lipopolysaccharide biosynthesis genes rfaK, rfaL,
RT rfaJ, and rfaZ of Escherichia coli K-12 and Salmonella typhimurium.";
RL J. BACTERIOL. 174:4746-4752(1992).
RN [2]

```

RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 94316500.

RA SOFIA H.J., BURLAND V., DANIELS D.L., PLUNKETT G. III, BLATTNER F.R.;

RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes.";

RL NUCLEIC ACIDS RES. 22:2576-2586(1994).

CC -1- FUNCTION: ADDS A N-ACETYL-D-GLUCOSAMINE GROUP ON A GLUCOSE
CC GROUP OF LPS.

CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE +
CC LIPOPOLYSACCHARIDE -> UDP + N-ACETYL-D-GLUCOSAMINYL-
CC LIPOPOLYSACCHARIDE.

CC -1- PATHWAY: LIPOPOLYSACCHARIDE CORE BIOSYNTHESIS.

CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE.

CC -1- SIMILARITY: SHOWS VERY LITTLE SIMILARITY TO S. TYPHIMURIUM RFAK.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

CC EMBL: M53398; G147573; -

CC DR EMBL: U00039; G466761; -

CC DR EMBL: AE000440; G1790053; -

CC DR PIR: C42981; C42981.

CC DR ECOGENE: EG1423; RFAK.

CC KW LIPOPOLYSACCHARIDE BIOSYNTHESIS; GLYCOSYLTRANSFERASE; TRANSFERASE.

CC FT CONFLICT 293 293 I -> M (IN REF. 1).

CC SQ SEQUENCE 357 AA; 41729 MW; 7DF52EFD CRC32;

alignment_scores:

157 GAAACCTTTATCAAAAAGTAAGTCGCCAAGAAGTCTAAGAATTAGT 208
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
496 GltuthValIyrcInIleketserIysGlnasnArgIasnGlnule 512
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
207 AGTGTTCCCATCACCCTGTGGAGTGTGCATTTCTAAAAAGATTTGATT 256
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
512 uillepealaserLeuIIeasnaspmethylsphe..... LeuIenSerp 527
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
257 CCGAAGATGACAT.....TATATTTAACCTTG 287
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
527 hetripIeaspngInglynsentTPYrGIuseerLeuylsIleLeuenth 543
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
286 GTGGGGGAANAAGTTATTAGAACACACAGTCTTCACCTTCGATACTGTAAA 335
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
544 IleasnashnIaspleuvValTyrylYstYserIleuIleLeuLeuEus 560
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

336 T 336
560 n 560

```

54  TCGATGATTCGAAATATGCTCAATTAATATTTATGCGGCAATTCCT 103
      ::::| |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
9   ValIleSerValIleuIleValThrLeuTyPheTyPheThrValglnrph 25
104 ACATTGCTTCGACAGATTAAATGCTGCGCCAAATTTTGATTTTAT 153
      ::::| |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
25  easnTyTrpPylsYsrArgsnValArgGlyProGluProValValrph 42
154 TTGAGACTTCTTATCAAAAGTA.....176
      ||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
42  heGlyAsnLeuLysAspSerAlaLeuArgLysLysAsnMetGlyValVal 58
177 .....ATGCTGCCAAGAA...GTCNAGGAATTAG 205
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
59  MetGluGluLeuTyPsrMetPheProGluIuLysValIleGlyIleTy 75

```

```

206 TAGGTGTCCTTCACCTGTTG.....GAG 231
75 rargnetlrserProcysleuValrArgspLeuGlValILllysh 92
232 GTGCTATTCTAAAGATTGAT...TTCCTGGAAAGACATTATATT 278
92 lsllelmetllysaephgluvalPheserapsrgrgllysluphe 108
279 AACCTTTGGTGGG.....GGAAANAGT 301
109 SerlysgluGlyleuGlylnslneupheHslAlaspolyAsphrr 125
302 TAGG.....ACCAAGCTCTCAGCTTCGACTTGTGA 125
125 pargthrlleuArgAsnArgpheThrrollephethrsergllysl 142
334 AATTAACTCTTTAT 347
142 yasnmetpethery 146
seq_name: SwissProt_37:TEGU_EBV
seq_documentation_block:
ID: TEGU_EBV STANDARD: PRT; 3149 AA.
AC P03186;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE LARGE TEGUMENT PROTEIN.
BN BLF1.
GN EPSTEIN-BARR VIRUS (STRAIN B95-8) (HUMAN HERPESVIRUS 4).
OS VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE; LYMPHOCRYPTOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 84270667.
RA BAR R., BANNIER A.T., BIGGIN M.D., DEININGER P.L., FARRELL P.J.,
RA GIBSON T.J., HAFEFUL G., HUDSON G.S., SATCHELL S.C., SEGUIN C.,
RA TUFFENELL P.S., BARRELL B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL NATURE 310:207-211(1984).
CC -1- FUNCTION: TEGUMENT PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EHV-1 24, EBV BLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC -----
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/).
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: V01555; G59094; -
DR PIR: A03747; Q0B88.
DR PIR: S32993; S32993.
DR SEQUENCE 3149 AA; 337954 MW; 8C0A19B4 CRC32;

```

```

alignment_scores:      length: 54
Quality: 58.00
Ratio: 1.706
Percent Similarity: 62.963
Percent Identity: 31.481
alignment_block:
US-09-030-606-223 x TEGU_EBV ..
Align seq 1/1 to: TEGU_EBV from: 1 to: 3149

```

204 AGTAGIGITCCMCACITGTTGGAGTGTGTAATTCTAAAGATTTTGA 253
|||:::||||::||| ::||| ||
682 SeraspIleProthrThr.....GluaspGluas 691

```

254 TTTCCTGGAAATGACAACTATATATTTTAA..TTGGGTGGGGGGAANNGT 3500
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
691 pmetheginsp.guvalp.hesera.nser.gluserysera 708
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
301 AATGAGACCAAGCTCTCACTCTGATACTGTAATTAATCTTTATTGC 3500
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
708 laprothnserprolethnleupnsphtralaarSergIntYTyrgln 7244
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
351 ACTGTGTTTGAC 362
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
725 TlnTlnrheasp 728

```

```

seq_name: SwissProt_37:UNG_BORBU
seq_documentation_block:
ID      UNG_BORBU          STANDARD;          PRT;          223 AA.
AC      O51082; Q44841;
DT      15-JUL-1998 (REL. 36, CREATED)
DT      15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT      15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE      URACIL-DNA GLYCOSYLASE (EC 3.2.2.-) (UDG).
GN      UNG OR UDG OR BB0053.
OS      BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; BORRELLIA.
OC      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-ATCC 35210 / B31;
RX      MEDLINE; 98065943.
RA      FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA      LAHIGHTA R., WHITE O., KETCHUM K.A., DODSON R., HICKLEY E.K., GWINN M.,
RA      DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
RA      PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALBERG S., HANSON M.,
RA      VAN VUUT R., PALMER N., ADAMS M.D., COCARYNE J.D., WEIDMAN J.,
RA      UTTERBACK T., WATHEY L., McDONALD L., ARTIACH P., Bowman C.,
RA      GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA      SMITH H.O., VENTER J.C.;
RT      "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT      burgdorferi".
RL      Nature 390:580-586(1997).
RN      [2]
RP      SEQUENCE OF 1-80 FROM N.A.
RC      STRAIN-ATCC 35210 / B31;
RA      GEEBIA J.A., BACKENSTROM P.B., ANDA P., COLEMAN J.L., BENNICH J.L.;
RA      SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC      -1- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE
CC      AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA
CC      POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its contents in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE001118; G2687922; -.
DR      EMBL; U57684; G1373351; -.
DR      TIGR; BB0053; -.
DR      PROSITE; PS00130; U_DNA_GLYCOSYLASE; 1.
DR      HSSP; P13051; IAKZ.
KW      DNA REPAIR; HYDROLASE; GLYCOSIDASE.
FT      ACT_SITE           67
FT      ACCT_SITE         67
FT      CONFLICT          32
FT      CONFLICT          73
FT      CONFLICT          79
SEQUENCE   223 AA; 25561 MW; EA56DOAE CRC32;
alignment_scores:
Quality:    57.50
Ratio:     1.127
Length:     97
Gaps:       4

```

Percent Similarity: 52.577 Percent Identity: 26.804

Alignment block:

US-09-030-606-223/rev x UNG_BORBU ..

Align seg 1/1 to: UNG_BORBU from: 1 to: 223

```

322 GAAGTGAAGACGTGTGCTATTAATCTTTCCCGACGACCAAGTAAATA 273
||| |||||
32 GLUTRylsThrlsLysasnlylsllePrlProPrlsleuIeph 48
272 TAATGT...CATTCAGGAAATCAAAATCTTTAGAAATGCA..... 233
||| |||||
48 easnIaphenSerleuProPrlsleuValIlellel 65
232 .....CACTCCAAGAGTGAAGGAGCACTACTAATCTT 197
||| |||||
65 lylGlnaspPrlTyrlsGlyLysasnlylslleuIaphen... 80
196 TAGACTTCTTGGCAGCATTTGATTAAGAGTCTCAAAATTAAT 147
||| |||||
81 .....SerValaspSerlysllelyslleProPrlsle 92
146 ACAAATTTGGCAGACATTTTAATCTTTCAGACAAATGTAAGAA 97
||| |||||
92 uGlnasnlylsllePrlsleuIeph...GlyLysSerleuIephThrl 108
96 TGGCCCAAAATATTAATGACCATTAATTCACAAATTC 56
||| |||||
108 lPrlasnlylslleuIephLysArgTPrIalIleGlnlyVal 121

```

seq_name: SwissProt_37:YCT2_BACFI

seq_documentation_block:

ID YCT2_BACFI STANDARD; PRT; 385 AA.

AC 004454;

DT 01-FEB-1994 (REL. 28, CREATED)

DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)

DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)

DE HYPOTHEtical 43.4 KD PROTEIN IN CTAF 3 REGION (ORF2).

OS BACILLUS FIRNUS.

OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;

OC BACILLUS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-OF4;

RX MEDLINE: 93107080.

RA OUIRK P.G., HICKS D.B., KRULWICH T.A.;

RT "Cloning of the cta operon from alkaliphilic Bacillus firmus OF4 and RT characterization of the pH-regulated cytochrome caa3 oxidase it encodes."

RL J. BIOL. CHEM. 268:678-685(1993).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announce/](http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch) or send an email to license@isb-sib.ch).

CC -----

CC EMBL: M94110; GI42791; -

CC HYPOTHEtical PROTEIN.

KW SEQUENCE 385 AA; 43410 MW; 98C98038 CRC32;

alignment_scores:

Quality: 57.50 Length: 110

Ratio: 1.009 Gaps: 5

Percent Similarity: 51.818 Percent Identity: 25.455

alignment_block:

US-09-030-606-223 x YCT2_BACFI ..

Align seg 1/1 to: YCT2_BACFI from: 1 to: 385

```

37 AAAAAATTAATCTTAGGACTGATATTGTAATTAAGTCAATTAATWTR 86
||| |||||
179 ArgasnlylslleGlnasp.....IleThrSerle 188
87 TTKTGCGCATTTCTTACATTTGCTTGTGACAAGATTAATGCTGTGCC 136
||| |||||
188 ulSerSerlePrlProGlyTyrLeuValThrPheLeuValTyrLeuIlea 205
137 AAAAAATTTGATTTATTATTGAGACT..... 162
||| |||||
205 lalLeuPheLeuPheMetLeuGluLeuProArgLeuArgLysleuTyr 221
163 TCTTATCAAAAGTAATGCTGCCAAAGAGCTTAAGAAATTAAGTGT 212
||| |||||
222 SerTyr.....LeuSerGluArgThrLysGlyLysValasnph 234
213 CCGMTC.....CTTGTTGAGAGTGTCTATTCFAAAGA 247
||| |||||
234 eMetThrSerArgLeuSerTyrValIleTrrpGlyPhePheLysalaglnp 251
248 TTTGATTTCTCGAATGACATTAATTTAATCTTGTGGGGAANA 297
||| |||||
251 heLeuValSer...IlellellePheIleValThrLeuIleGlyleuLeu 266
298 GTTATAGACACACAGCTCTCACTTGATTA 327
||| |||||
267 PheIlealPrlProGluValalaleuLeuMet 276

```

seq_name: SwissProt_37:GHR_HUMAN

seq_documentation_block:

ID GHR_HUMAN STANDARD; PRT; 638 AA.

AC P10912;

DT 01-JUL-1989 (REL. 11, CREATED)

DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE GROWTH HORMONE RECEPTOR PRECURSOR (GH RECEPTOR) (SERUM BINDING DE PROTEIN).

GN GHR.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE-LIVER.

RX MEDLINE: 88065896.

RA LEUNG D.W., SPENCER S.A., CACHIANES G., HAMMONDS R.G., COLLINS C., HENZEL W.J., BARNARD R., WATERS M.J., WOOD W.I.;

RT "Growth hormone receptor and serum binding protein: purification, RT cloning and expression."

RL NATURE 330:537-543(1987).

CC [2]

CC SEQUENCE FROM N.A.

CC MEDLINE: 90046742.

CC GODOWSKI P.J., LEUNG D.W., MEACHAM L.R., GALGANI J.P., HELLMITSS R., KREER R., ROTWEIN P.S., PARKS J.S., LARON Z., WOOD W.I.;

CC "Characterization of the human growth hormone receptor gene and RT demonstration of a partial gene deletion in two patients with Laron-type dwarfism."

CC PROC. NATL. ACAD. SCI. U.S.A. 86:8083-8087(1989).

RL [3]

RP DISULFIDE BONDS.

RX MEDLINE: 90153957.

RA FUH G., MULKERLIN M.G., BASS S., MCFARLAND N., BROCHIER M., BOURREL J.H., LIGHT D.R., WELLS J.A.;

RT "The human growth hormone receptor. Secretion from Escherichia coli and disulfide bonding pattern of the extracellular binding domain."

RL J. BIOL. CHEM. 265:3111-3115(1990).

RN [4]

RP VARIANT LARON DWARFISM SER-114.

RX MEDLINE: 89384829.
 RA AMSELEM S., DUQUESNOY P., ATTREE O., NOVELLI G., BOUSNINA S.,
 RA POSTELVAINAY M.-C., GOOSSENS M.;
 RT "Laron dwarfism and mutations of the growth hormone-receptor gene";
 RL NEW ENGL. J. MED. 321:989-995(1989).
 RN [5]
 RP VARIANTS LARON DWARFISM.
 RX MEDLINE: 93278381.
 RA AMSELEM S., DUQUESNOY P., DURIEZ B., DASTOT F., SORBIER M.-L.,
 RA VALLEIX S., GOOSSENS M.;
 RT "Spectrum of growth hormone receptor mutations and associated
 RT haplotypes in Laron syndrome";
 RL HUM. MOL. GENET. 2:355-359(1993).
 RN [6]
 RP VARIANT LARON DWARFISM HIS-170.
 RX MEDLINE: 94185645.
 RA DUQUESNOY P., SORBIER M.-L., DURIEZ B., DASTOT F., BUCHANAN C.R.,
 RA SAVAGE M.O., PREECE M.A., CRAESCU C.T., BLOUDUIT Y., GOOSSENS M.,
 RA AMSELEM S.;
 RT "A single amino acid substitution in the exoplasmic domain of the
 RT human growth hormone (GH) receptor confers familial GH resistance
 RT (Laron syndrome) with positive GH-binding activity by abolishing
 RT receptor homodimerization";
 RL EMBO J. 13:1386-1395(1994).
 RN [7]
 RP VARIANTS IDIOPATHIC SHORT STATURE LYS-62; CYS-179 AND ASP-242.
 RX MEDLINE: 96013502.
 RA GODDARD A.D., COVELLO R., LIUO S.-M., CLACKSON T., ATTIE K.M.,
 RA GESUNDHEIT N., RUNDLE A.C., WELLS J.A., CARLSON L.M.S.;
 RT "Mutations of the growth hormone receptor in children with idiopathic
 RT short stature. The Growth Hormone Insensitivity Study Group";
 RL NEW ENGL. J. MED. 333:1093-1098(1995).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 19-256.
 RX MEDLINE: 92196577.
 RA DE VOS A.M., ULTSCH M., KOSSIAKOFF A.A.;
 RT "Human growth hormone and extracellular domain of its receptor:
 RT crystal structure of the complex";
 RL SCIENCE 255:306-312(1992).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 19-256.
 RX MEDLINE: 97113023.
 RA SUNDSTROM M., LUNDQVIST T., RODIN J., GIEBEL L.B., MILLIGAN D.,
 RA NORSTEDT G.;
 RT "Crystal structure of an antagonist mutant of human growth hormone,
 RT G120R, in complex with its receptor at 2.9-A resolution";
 RL J. BIOL. CHEM. 271:32197-32203(1996).
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR PITUITARY GLAND GROWTH HORMONE.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- DISEASE: DEFICIENCY IN GHR IS THE CAUSE OF PITUITARY DWARFISM II
 CC (LARON-TYPE PITUITARY DWARFISM OR LARON SYNDROME (LS)). IT ALSO
 CC CAUSES IDIOPATHIC SHORT STATURE.
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X06562: G31738: -
 DR EMBL: M28466: G183171: -
 DR EMBL: M28456: G183171: JOINED.
 DR EMBL: M28459: G183171: JOINED.
 DR EMBL: M28460: G183171: JOINED.
 DR EMBL: M28461: G183171: JOINED.
 DR EMBL: M28462: G183171: JOINED.
 DR EMBL: M28463: G183171: JOINED.
 DR EMBL: M28464: G183171: JOINED.

DR EMBL: M28465: G183171: JOINED.
 DR PIR: S04530: S04530.
 DR PIR: A33991: A33991.
 DR PDB: 3HHR: 30-APR-94.
 DR PDB: 1HWG: 19-NOV-97.
 DR PDB: 1HWI: 19-NOV-97.
 DR PDB: 1AXI: 28-JAN-98.
 DR PDB: 1A22: 29-APR-98.
 DR MIM: 600946: -
 DR MIM: 262500: -
 DR PROSITE: PS00241: RECEPTOR_CYTOKINES_1; 1.
 DR PROSITE: PS00340: RECEPTOR_CYTOKINES_2; FALSE_NEG.
 DR PFM: PF00041: fn3; 1.
 KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL; 3D-STRUCTURE;
 KW DWARFISM; DISEASE MUTATION.
 FT SIGNAL 1 18
 FT CHAIN 19 638
 FT DOMAIN 19 264
 FT TRANSMEM 265 288
 FT DOMAIN 289 638
 FT DOMAIN 145 252
 FT DISULFID 56 66
 FT DISULFID 101 112
 FT DISULFID 126 140
 FT CARBOHYD 46 46
 FT CARBOHYD 115 115
 FT CARBOHYD 156 156
 FT CARBOHYD 161 161
 FT CARBOHYD 200 200
 FT VARIANT 62 62
 FT VARIANT 89 89
 FT VARIANT 114 114
 FT VARIANT 143 143
 FT VARIANT 162 162
 FT VARIANT 170 170
 FT VARIANT 179 179
 FT VARIANT 229 229
 FT VARIANT 242 242
 FT CONFLICT 344 344
 FT STRAND 53 58
 FT STRAND 64 68
 FT STRAND 82 88
 FT STRAND 99 100
 FT STRAND 104 107
 FT TURN 109 110
 FT STRAND 111 114
 FT TURN 116 117
 FT STRAND 124 131
 FT TURN 132 133
 FT STRAND 134 142
 FT HELIX 143 145
 FT STRAND 147 147
 FT STRAND 153 162
 FT TURN 164 165
 FT STRAND 168 176
 FT TURN 179 180
 FT TURN 183 186
 FT STRAND 190 198
 FT TURN 200 200
 FT STRAND 205 206
 FT STRAND 210 210
 FT STRAND 214 221
 FT TURN 222 223
 FT STRAND 225 234
 FT STRAND 247 250
 SQ SEQUENCE 638 AA: 71499 MW: 5692540C CRC32;
 alignment_scores: 57.50 Length: 53
 Quality: 1.983 Gaps: 2
 Ratio: 1.983

Percent Similarity: 54.717 Percent Identity: 30.189

alignment_block:

US-09-030-606-223 x GHR_HUMAN ..

Align seg 1/1 to: GHR_HUMAN from: 1 to: 638

```
11  ACAAACAAACAAATCTCATTCAGAAAAATATCTAGGAGACTGATAT 60
   ::|||:::|||||
232 SerLysGlnArgAsnSer..... 237
61  TGGTAATTATGTCGAATTAAATWRTTRTKTGGGCGCATTCCTTACATTGT 110
   |||||:::|||||:::|||||
238 .GlyAsnTyrGlyGlnPheSerGluValLeu.....TyrValThrLeuP 252
111 CTTGACAAGATTAAATGTCCTGCGCAAAATTGTGATTTATTGGAGA 160
   :::|||||:::|||||
252 roGlnMetSerGlnPheThrCysGlnGluAspPheTyrPheProTyrPleu 268
161 CTCTCTATC 169
   |||:::|||||
269 LeuIleIle 271
```

This Page Blank (uspto)

DR PFAM: PF00083; sugar_tr; 1.
SQ SEQUENCE 586 AA; 64272 MW; 3E637F95 CRC32;

alignment_scores:
Quality: 68.00 Length: 92
Ratio: 1.581 Gaps: 5
Percent Similarity: 46.739 Percent Identity: 26.087

alignment_block:
US-09-030-606-223 x Q07824 ..

Align seg 1/1 to: Q07824 from: 1 to: 586

```

91 GGGGCAATTCCTTCATGCTTCGACAAGATAAATGTCGCCAAA 140
||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
414 GlyIuLeuProTyrIleAlaIleIleGlyMetValCysAlaI 430
141 TTTGTAATTTATTTGAGACTTCTATCAAAAGTAATGCTCCAAA 190
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
430 aPheIleTyrPheMetAspPheAspTyrLeuLysArgCys....Arg 445
191 AGTTAAGAAATTAGTA..... 207
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
445 IalysGlyLysLeuValProGluAlaArgLeuTyrAlaMetValIleAla 461
208 .....GTGTTCCM...TCACCTGTTGGAGTGTGCTATCTAAAGAT 248
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
462 GlyThrValPheProIleGlyIleLeuTyr..... 471
249 TTGATTTCTCGAATGACATTA.....ATTGTA 280
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
472 ....PheCysTyrPheGlyTyrTyrProHisLysIleHisTyrMetValP 487
281 CTTTGCTGGGGGAANAAGTTTAGCA 306
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
487 roThrValGlyAlaPheIleGly 495

```

seq_name: sp_virus:039250

seq_documentation_block:
ID 039250 PRELIMINARY; PRT; 1080 AA.

AC 039250;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE COUNTERPART OF HSV-1 GENE US2 AND VZV GENE 6.

GN 7.
OS Equine herpesvirus 4.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirinae.
RN [1]
RP SEQUENCE OF 1-198 FROM N.A.
RC STRAIN-NS80567;
RC MEDLINE: 94058670.

RA RIGGIO M.P., ONIONS D.E.;
RT "DNA sequence of a gene cluster in the equine herpesvirus-4 genome
RT which contains a newly identified herpesvirus gene encoding a
RT membrane protein.";
RT Arch. Virol. 133:171-178(1993).

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-NS80567;
RC MEDLINE: 98264497.
RA TELFORD E.A.R., WATSON M.S., PERRY J., CULLINANE A.A., DAVISON A.J.;
RT "The DNA sequence of equine herpesvirus-4.";
RT J. Gen. Virol. 79:1197-1203(1998).

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-NS80567;
RA TELFORD E.A., WATSON M.S., PERRY J., CULLINANE A.A., DAVISON A.J.;
RT Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF030027; AAC59589.1; -

SQ SEQUENCE 1080 AA; 120603 MW; 2A8793CB CRC32;

alignment_scores:
Quality: 66.00 Length: 125
Ratio: 1.179 Gaps: 8
Percent Similarity: 44.800 Percent Identity: 26.400

alignment_block:
US-09-030-606-223 x 039250 ..

Align seg 1/1 to: 039250 from: 1 to: 1080

```

3 ACAACAACAACAAAAAACAATTCATTCAGAAAAAATATCTTAGG 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
773 AsnLysThrThrAspLysThrProLeuGln..... 782
53 ACTGAAATTCGTAATATGTCATTTAAATWRTTKTGCGGCATTTC 102
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
783 .....LeuLeuLeuGlnThrAsn.....AlaC 790
103 TACATTCCTTGACAG.....ATTAATGCTGTGCCAAA 140
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
790 ySgIncysThrGlyLysMetGlyPheArgIleThrValProValProPro 806
141 TTTGTAATTTATTTGAGACTTCTATCAAAAGTAATGCTGCCAAGA 190
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
807 ProTyrIleLeuAlaGly.....ProGluAl 815
191 AGCTAAGAAATTAGTAGTGT.....CCCMCACTGTTTGAG.... 230
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
815 aLeuArgGlyValAlaArgIleIleGlnGlnAlaValIleGlnLarg 832
231 .....TGCTATTCCTAAACATTTGCTTCCTCGAA 263
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
832 hPheThrGluSerMetCysSerValLeuArgAspPheSerPheLeuAsp 848
264 TGACAATATATTTTAACTTGTGGGGGAANAAGTATATGACACAGT 313
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
849 ThrGlyValTyrSerHis.....GlyArgSerLeuArgLeuProPh 862
314 CTCACCTTCGATACCTGTAATTA 338
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
862 ePhe.....CysLysVal 866

```

seq_name: sp_bacteria:Q56394

seq_documentation_block:
ID Q56394 PRELIMINARY; PRT; 301 AA.

AC Q56394;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, last annotation update)
DE XYNB.

GN XYNB.
OS Thermoanaerobacter thermosulfurogenes (Clostridium
OC thermosulfurogenes).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Thermoanaerobacter
OC group; Thermoanaerobacterium.

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EMI (DSM 3896);
RA MATSCHER M., SAHM K., BAHL H.;
RT Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U50952; AAB08045.1; -
DR PFAM: PF00528; BPD_transp; 1.
DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBER; 1.
KW Transport; Transmembrane.

SQ SEQUENCE 301 AA; 33755 MW; 6630324E CRC32;

alignment_scores:
Quality: 66.00 Length: 108

Ratio: 1.200 Gaps: 4
Percent Similarity: 50.926 Percent Identity: 24.074

alignment_block:
US-09-030-606-223 x 056394 ..

Align seg 1/1 to: 056394 from: 1 to: 301

```

70 TGCATATTAATTTATTTTGGGCGCATTTCTTACATTGCTGTGCAAG 119
   ::::::::::::::::::::|::|::|::|::|::|::|::|::|::|
121 TYSERMETYRPHETHRSERILEUPEPROTYRILEULEULEAR 137
   ::::::::::::::::::::|::|::|::|::|::|::|::|::|::|
120 ATTAAGATGCTGTGCAAAATTTGTATTTATTTGAGACTTCTATC 169
   ||| ::::::::::::::|::|::|::|::|::|::|::|::|::|
137 GASPEUHLISLEUMETASNAPHLEUVALTYR..... 148
   ||| ::::::::::::::|::|::|::|::|::|::|::|::|::|
170 AAAAGTAATGCTGCCAAGAGAGCTAAGAAATAGTAGTCCCTCA 219
   ::::::::::::::|::|::|::|::|::|::|::|::|::|
149 ..... 152
220 CTGTCT.....TGAGTGTGCTATTTAAAGATTGATTTCTGTGCA 263
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
153 ILEVALSERALATRPASNILEMET.....VALVALTYSERTYRIL 166
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
264 TGACATATTAATTTTAACTTTGTGGGGAANAGTATAGACCAAGT 313
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|
166 CASPOLYLEUPROALASERLEUVALGLUSERALALYSTLEASGLYALAS 183
314 CTTCACCTCTGTACTTGTAAATTAATCTTT.....ATTCACCTT 354
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|
183 ERLULEUARGILELLEPHERSERILELLEPHEPROLEUSERVALPROVAL 199
355 GTTTGACCATTAACCTATATCTT 378
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
200 LEUALATRIHLEHLEUPHEVAL 207

```

seq_name: sp_virus:090699

seq_documentation_block:

```

ID 090699 PRELIMINARY: PRT: 1818 AA.
AC 090699;
DT 01-NOV-1998 (TREMBLREL. 08, Created)
DT 01-NOV-1998 (TREMBLREL. 08, Last sequence update)
DT 01-MAY-1999 (TREMBLREL. 10, Last annotation update)
DE HYPOTHETICAL 207.4 KD PROTEIN.
OS Beet virus Q.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
RN [1]
RX MEDLINE: 98378056.
RA KOENIG R., FLEIS C., BEIER C., COMMANDEUR U.;
RT "Genome properties of beet virus Q, a new furo-like virus from
RT sugarbeet, determined from unpurified virus."
RL J. Gen. Virol. 79:2027-2036(1998).
DR EMBL: AJ223596; CA11457.1; -.
DR PRAM: PF00978; RNA_dep_RNApol2; 1.
KW Hypothetical protein.
SQ SEQUENCE 1818 AA; 207431 MW; 7003FEEF CRC32;

```

alignment_scores:

Quality: 65.50 Length: 95
Ratio: 1.394 Gaps: 5
Percent Similarity: 49.474 Percent Identity: 25.263

alignment_block:

US-09-030-606-223 x 090699 ..

Align seg 1/1 to: 090699 from: 1 to: 1818

```

90 TGGGCAATTCCTTACATTTGCTGACAGATAAAGTCTGTGCCAA 139
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|
1647 TRPSERIALALALEUSERLEUIGLUSERLEUPROLEUGLALYALALR 1664

```

seq_name: sp_virus:056790

seq_documentation_block:

```

ID 056790 PRELIMINARY: PRT: 1778 AA.
AC 056790;
DT 01-JUN-1998 (TREMBLREL. 06, Created)
DT 01-JUN-1998 (TREMBLREL. 06, Last sequence update)
DT 01-MAY-1999 (TREMBLREL. 10, Last annotation update)
DE 204 KDA PROTEIN.
OS Beet soil-borne virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Furovirus.
RN [1]
RX MEDLINE: 98062137.
RA KOENIG R., LOSS S.;
RT "Beet soil-borne virus RNA 1: genetic analysis enabled by a starting
RT sequence generated with primers to highly conserved helicase-encoding
RT domains."
RL J. Gen. Virol. 78:3161-3165(1997).
DR EMBL: Z97873; CAB10764.1; -.
DR PRAM: PF00978; RNA_dep_RNApol2; 1.
DR PRAM: PF00149; STphosphatase; 1.
SQ SEQUENCE 1778 AA; 203690 MW; BE65D96F CRC32;

```

alignment_scores:

Quality: 64.50 Length: 95
Ratio: 1.372 Gaps: 5
Percent Similarity: 49.474 Percent Identity: 27.368

alignment_block:

US-09-030-606-223 x 056790 ..

Align seg 1/1 to: 056790 from: 1 to: 1778

```

90 TGGGCAATTCCTTACATTTGCTGACAGATAAAGTCTGTGCCAA 139
   ||::|::|::|::|::|::|::|::|::|::|::|::|
1607 TRPSERIALALALEUVALLEUVALSPALALEUPROLEUGLALYALALY 1623
140 ATTTGTATTTATTTATTTGAGACTTCTATCAAAAGTATGTCGCAAG 189
   ||::|::|::|::|::|::|::|::|::|::|::|::|
1623 SPHEMETVALPHEGLYGLYASPA...SERLEUVALPHEPHEPROLYSA 1639
190 AAGCTAAGCAATTAGTAGTGTCTCCMTCACTGTGTGGAGTGTCTATT 239
   ::::|::|::|::|::|::|::|::|::|::|::|::|
1639 SNET.....ASNLEUALASPPROCYSGLYARGLEUALASER 1651
240 CTAAGATTTGATTTCTGTGGAATGACAAATTAATTTTAACTTTGTGTG 289
   ||| ::|::|::|::|::|::|::|::|::|::|::|::|
1652 LEUTRIPASNPHASPCYS.....LYSPHERHASNPHGLINAS 1664

```

290 G.....GGAANAGTATAGGA 306
 1664 nasmetPheCysGlySPhenLeuLysILeGlyGlnsnrtyrlys. 1680
 307 CCACAGCTCTTCACTTGTGATACCTGTAATATATC 341
 1681PheAlaProAspProPheLysLeu 1689

seq_name: sp.bacteria:084258

seq_documentation_block: PRT; 414 AA.
 ID 084258 PRELIMINARY;
 AC 084258;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
 DE 01-NOV-1998 (TREMBLrel. 08, last annotation update)
 DE HYPOTHETICAL 45.8 KD PROTEIN.
 GN CT256.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-D/UM-3/CX;
 RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
 RA MITCHELL W.P., OLINGER L., TATISOV R.L., ZHAO Q., KOONIN E.V.,
 RA DAVIS R.W.;
 RT "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
 RT Chlamydia trachomatis."
 RL Science 0:0-0(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-D/UM-3/CX;
 RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
 RA MITCHELL W.P., OLINGER L., TATISOV R.L., ZHAO Q., KOONIN E.V.,
 RA DAVIS R.W.;
 RT Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AE001298; AAC67849.1;
 DR Hypothetical protein.
 KW SEQUENCE 414 AA; 45838 MW; 0B5C5CAD CRC32;

alignment_scores: Quality: 64.50 Length: 104
 Ratio: 1.075 Gaps: 6
 Percent Similarity: 57.692 Percent Identity: 29.808

alignment_block:

US-09-030-606-223 x 084258 ..

Align seg 1/1 to: 084258 from: 1 to: 414

106 ATGTCTGTGACAGATTAAATGCTGTGCCAAATTTTGA..... 147
 7 ValIleuThrAlaIlePheValIeuCysSerGlyPheValSerLeu 23
 148TTTATTGTGAGACT.....TCTT 166
 23 rHisIleAlaLeuPheSerLeuProSerSerLeuIleAlaIstYserH 40
 167 ATCAAAAGTAATGCTGCCAAGAAGTCTAAGATTAGTAGTTCCTCCM 216
 40 IsSerLysAsnArgGlnLeuArgGlnIleAlaAsnLeuMetAlaIstPro 56
 217 TCA...CTGTGTGAGTGTCTATTTCTAAAGATTGATTCTCTGGA 263
 57 AsnHisIleuLeuMetThrLeuValPhePheAspIleGlyIleAsnIleG 73
 264 T...GACATATATATTTAACTTGTGGGGGGAANAGTTAGACAC 310
 73 yValGlnAsnCysIleAlaThrLeuValGlyAsp.....S 85
 311 AGTCTTCACTTGTGATACCTGTAATTAATCTTATTGACCTGTTTG 360

85 eraIaSerLeuLeuThrValGly.....ValProIleAlaLeu 98
 361 ACCATTAAGCTA 372
 99 ThrLeuValLeu 102

seq_name: sp.plant:004582

seq_documentation_block: PRT; 559 AA.
 ID 004582 PRELIMINARY;
 AC 004582;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, last sequence update)
 DE 01-MAY-1999 (TREMBLrel. 10, last annotation update)
 DE SIMILARITY TO TRITICUM AESTIVUM SERPIN.
 GN F19K23.10.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA OSBORNE B.I., VYOTSKAIA V.S., TORIUMI M., YU G., OJI O., SHEN Y.K.,
 RA ARAUJO R., AU M., BUEHLER E., CONWAY A.B., CONWAY A.R., DEMAR K.,
 RA PENG J., KIM C., KURZ D., LI Y., SHINN P., SUN H., DAVIS R.W.,
 RA ECKER J.R., FEDERSPIEL N.A., THEOLOGIS A.;
 RT Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AC000375; AAB60763.1;
 DR MENDEL; 16823; ARATH;1514;ml16823.
 DR PFAM: PF00079; serpin; 4.
 DR PROSITE: PS00264; SERPIN; 2.
 KW Serpin.
 KW SEQUENCE 559 AA; 62722 MW; B8CC961A CRC32;

alignment_scores: Quality: 63.50 Length: 81
 Ratio: 1.549 Gaps: 3
 Percent Similarity: 50.617 Percent Identity: 28.395

alignment_block:

US-09-030-606-223 x 004582 ..

Align seg 1/1 to: 004582 from: 1 to: 559

123 AAATGTCTGTGCCAAATTTGTATTTATTGGA..... 158
 467 LysPheArgIleProLysPheLysIleGlnPheGlyPheGlnIaIaSer 483
 159GACTTCTATCAAAAGTAATGC 180
 483 rAlaPheSerAspPheGlnLeuAspValSerPheYrGlnIstYThrLeu 500
 500 IeGlnIleAspGlnLysGlyThrGlnIaIaValThrPheThrAlaPheArg 516
 225TTGAGTGTGCTATTTCTAAAGATTGATTCTCTGGAAG 265
 517 SerAlaIstYrLeuGlyCysAlaLeuValIstProIleAspPheValAla.A 533
 266 ACAATTATATTTAACTTGTGGGGGGAANAGTTATAGGA 306
 533 sPHisProPheLeuPheLeuIleArgGlnIstYrGlnIstYrGly 546

seq_name: sp.bacteria:Q59319
 seq_documentation_block: PRT; 825 AA.
 ID Q59319 PRELIMINARY;
 AC Q59319;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)


```

      |||  ::|||::|||  |||::  ::  |||::
57 PROHISGLYArgLeuGIUTyrAlaGlnHisGlnGlyArgIleLysAsnAl 73
242 TAGAATAGCAGACACTCCAAACAAGTGAAGGAGACACTACTAATTCCTTAGA 193
      |||  ||| ||| ||| ||| ::
73 aATgGUAAlAHISserGln..... 79
192 CTTCCTTTGGCAGCATTACTTTGATAGAGAGTCTCCAAATAAATACAA 143
      ::|||::|||::  ::|||::|||::
80 ..... 89
142 AATTGTGCAGACACATTTTAATCTTGTCAGACAAATGTAGGAAATGCC 93
      ::|||  ::|||  |||  |||::  |||
90 SerPheIleAspIleuAlaSerLeuValProThrCysAsnAlaMetSe 106
92 CCMAAYAVWATTAATGACCATTAAT 65
      ::  ||| ||| ||| ||| ::
106 rArgLysLeuAspLysLeuThrValLeu 115

```

```

92 CCMAAYAVWATTAATGACCATTAAT 65
      ::  ||| ||| ||| ||| ::
106 rArgLysLeuAspLysLeuThrValLeu 115

```

seq_name: sp_rodent:088810

seq_documentation_block:

```

ID 088810 PRELIMINARY; PRT; 626 AA.
AC 088810;
DT 01-NOV-1998 (TRENBLREL. 08, Created)
DT 01-NOV-1998 (TRENBLREL. 08, Last sequence update)
DT 01-MAY-1999 (TRENBLREL. 10, Last annotation update)
DE BMAL1B.
GN BMAL1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 98407895.
RA HONMA S., IKEDA M., ABE H., TANAHASHI Y., NAMIHARA M., HONMA K.,
RA NOMURA M.;
RT "Circadian oscillation of BMAL1, a partner of a mammalian clock gene
RT clock, in rat suprachiasmatic nucleus."
RL Biochem. Biophys. Res. Commun. 250:83-87(1998).
DR EMBL; AB012600; BAA33450.1; -.
DR PFAM; PF00010; HLH; 1.
DR PFAM; PF00989; PAS; 1.
SQ SEQUENCE 626 AA; 68531 MW; 60BF0E7D CRC32;

```

alignment_scores:

```

      Quality: 62.50      Length: 76
      Ratio: 1.689      Gaps: 1
Percent Similarity: 48.684 Percent Identity: 23.684

```

alignment_block:

US-09-030-606-223/rev x 088810 ..

Align seg 1/1 to: 088810 from: 1 to: 626

```

292 CCCCCCAAGTTAAATATTAATTCATTCGAGAAATCAAAATCTTT 243
      |||  ::|||::|||  |||::  ::  |||::
57 PROHISGLYArgLeuGIUTyrAlaGlnHisGlnGlyArgIleLysAsnAl 73
242 TAGAATAGCAGACACTCCAAACAAGTGAAGGAGACACTACTAATTCCTTAGA 193
      |||  ||| ||| ||| ||| ::
73 aATgGUAAlAHISserGln..... 79
192 CTTCCTTTGGCAGCATTACTTTGATAGAGAGTCTCCAAATAAATACAA 143
      ::|||::|||::  ::|||::|||::
80 ..... 89
142 AATTGTGCAGACACATTTTAATCTTGTCAGACAAATGTAGGAAATGCC 93
      ::|||  ::|||  |||  |||::  |||
90 SerPheIleAspIleuAlaSerLeuValProThrCysAsnAlaMetSe 106

```


OM nucleic - nucleic search, using sw model

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result
No.

27

3. 6

5

1:	em.est1.*
2:	em.est2.*
3:	em.est3.*
4:	em.est4.*
5:	em.est5.*
6:	em.est6.*
7:	em.est7.*
8:	em.est8.*
9:	em.est9.*
10:	em.est10.*
11:	em.est11.*
12:	em.est12.*
13:	em.est13.*
14:	em.est14.*
15:	em.est15.*
16:	em.est16.*
17:	em.est17.*
18:	em.est18.*
19:	em.est19.*
20:	gb.est1.*
21:	gb.est2.*
22:	gb.est3.*
23:	gb.est4.*
24:	gb.est5.*
25:	gb.est6.*
26:	gb.est7.*
27:	gb.est8.*
28:	gb.est9.*
29:	gb.est10.*
30:	gb.est11.*
31:	gb.est12.*
32:	gb.est13.*
33:	gb.est14.*
34:	gb.est15.*
35:	gb.est16.*
36:	gb.est17.*
37:	gb.est18.*
38:	gb.est19.*
39:	gb.est20.*
40:	gb.est21.*
41:	gb.est22.*
42:	gb.est23.*
43:	gb.est24.*
44:	gb.est25.*
45:	gb.est26.*
46:	gb.est27.*
47:	gb.est28.*
48:	gb.est29.*
49:	gb.est30.*
50:	gb.est31.*
51:	gb.est32.*
52:	em.est30.*

```
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result				Description
No.	Score	Query	ID	
1	376.8	98.4	400	AA653016
2	374	97.7	387	AA579247
3	369.8	96.6	522	W56075
4	347.2	90.7	469	AI188401
5	339.4	88.6	432	C75566
6	334.4	87.3	472	AI027556
7	327.4	85.5	389	AA879150
8	320.6	83.9	442	AI807043
9	317.4	82.9	444	AI522136
10	313.4	81.8	419	C75584
11	303.4	79.2	496	AI005274
12	291.6	76.1	473	AI051146
13	288	75.2	407	AA622892
14	280.4	73.2	344	C14656
15	277	72.3	463	AI420653
16	261.2	68.2	324	AI318914
17	257	67.1	442	AI382141
18	229.6	59.9	231	AA653045
19	227	59.3	268	AA093399
20	213	55.6	225	AA652361
21	203	53.0	332	AA657851
22	189.6	49.5	352	AA093469
23	185.4	48.4	246	AA603128
24	185	48.3	372	AI214972
25	181.4	47.4	200	AA372885
26	171	44.6	373	AA560097
27	158.4	41.4	282	AA216139
28	151.4	39.5	255	H83672
29	130.8	34.2	410	AI16199
30	118.2	30.9	703	AI56846
31	102.8	26.8	110	AA503672
32	101.2	26.4	309	AI761860
33	92.4	24.1	576	C85936
34	92.4	24.1	438	AI133916
35	92.4	24.1	373	AI450559
36	91	23.8	425	AA137503
37	83.2	21.7	231	AA658025
38	76	19.8	275	AV003262
39	68.6	17.9	210	AV010724
40	64	16.7	298	AV044015
41	62.4	16.3	194	AI350423
42	42.6	11.1	427	AI152967
43	42.6	11.1	437	AI438145
44	41.2	10.8	412	AI289102
45	40.8	10.7	392	D61957
			23	D61957 HUM26F10B
				AA653016 n68901.s
				AA579247 n136d10.s
				W56075 zc56b05.r1
				AI188401 qd08e02.x
				C75566 C75566 Huma
				AI027556 ow52c09.x
				AA879150 n84e05.s
				AI166743 qd89d03.x
				AI522136 t147f03.x
				C75584 C75584 Huma
				AI005274 ou07f12.x
				AI051146 oy49d04.x
				AA622892 np58c12.s
				C14656 C14656 Clon
				AI420653 t13b04.x
				AA318914 ESR21101
				AI382141 te30c04.x
				AA653045 ns71f07.s
				AA093399 c11544.se
				AA652361 ns73d07.s
				AA657851 nu08d07.s
				AA093469 kK5000.se
				AA603128 tr28h07.r
				AI214972 qm31f09.x
				AA372885 ESR84852
				AA560097 nh51n04.s
				AA216139 hp0520.se
				H83672 yv84b03.s1
				AI16199 yb96e01.s1
				AI56846 pN2.1.07-
				AA503672 nq78h10.s
				AI761860 yb96e01.r1
				C85936 C85936 Mous
				AI133916 uJ38g06.x
				AI450559 mqt79f08.x
				AA137503 mqt79f07.r
				AA658025 nu15d01.s
				AV003262 AV003262
				AV010724 AV010724
				AV044015 AV044015
				AI350423 q17d05.x
				AI152967 u654d06.r
				AI438145 m145c02.x
				AI289102 qw26f10B

ALIGNMENTS

RESULT	1
AA653016	
LOCUS	
DEFINITION	
ACCESSION	
NID	
VERSION	
AA653016	400 bp mRNA
AA653016	nc88901.s1 NCI_CGAP_Pr2 Homo sapiens CDNA clone IMAGE:1188816, mRNA
AA653016	sequence.
AA653016	
92584668	
AA653016.1	GI:2584668

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 400)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1325236.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuquai,
M.D., Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Kitzman, Ph.D.
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html

Insert Length: 678 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 376.

FEATURES
Location/Qualifiers
1..400
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="118816"
/clone_lib="NCI_CGAP_P12"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: PAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected preneoplastic cells
histologically-determined to be prostatic intraepithelial
neoplasia 2 (PIN2) cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into PAMP10 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Kitzman."

BASE COUNT 130 a 55 c 70 g 145 t

ORIGIN

Query Match 98.4%; Score 376.8; DB 36; Length 400;
Best Local Similarity 97.9%; Pred. No. 3.4e-68;
Matches 375; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

1 AAAACAACAACAACAACAACAACAATTCATCAGAAAAATATCTTAGGACTGATAT 60
DB AAAACAACAACAACAACAACAACAATTCATCAGAAAAATATCTTAGGACTGATAT 73
61 TGGTAATTAAGTCAATTAATWRTTCTGGGCAATTCCTTACATGCTTGACAGA 120
DB TGGTAATTAAGTCAATTAATWRTTCTGGGCAATTCCTTACATGCTTGACAGA 133
74 TGGTAATTAAGTCAATTAATWRTTCTGGGCAATTCCTTACATGCTTGACAGA 133
121 TTTAAATGCTGTGCCAAATTTTGTATTTATTTGGAGACTCTTATCAAAAGTAATGC 180
DB TTTAAATGCTGTGCCAAATTTTGTATTTATTTGGAGACTCTTATCAAAAGTAATGC 193
134 TTTAAATGCTGTGCCAAATTTTGTATTTATTTGGAGACTCTTATCAAAAGTAATGC 193
181 TGGCAAGAGAGCTAAGGATAGTAGTGCCATCTGTTGGAGTGGCATTC 240
DB TGGCAAGAGAGCTAAGGATAGTAGTGCCATCTGTTGGAGTGGCATTC 253
194 TGGCAAGAGAGCTAAGGATAGTAGTGCCATCTGTTGGAGTGGCATTC 253
241 TAAAGCTTTTGTATTCCTGGAATGACAAATATATTTTAACCTTGGTGGGGAANAAGTT 300

DB 254 TAAAGCTTTTGTATTCCTGGAATGACAAATATATTTTAACCTTGGTGGGGAANAAGTT 313
QY 301 ATAGACCCAGCTCTTCTGATCTGATTAATTAATCTTTATGCACTGTTTG 360
DB 314 ATAGACCCAGCTCTTCTGATCTGATTAATTAATCTTTATGCACTGTTTG 373
QY 361 ACCATTACCTATATGTTTAAA 383
DB 374 ACCATTACCTATATGTTTAAA 396

RESULT 2
AA579247 387 bp mRNA EST 12-SEP-1997
LOCUS n136d10.s1 NCI_CGAP_P12 Homo sapiens cDNA clone IMAGE:915859, mRNA
DEFINITION sequence.
ACCESSION AA579247
NID 92357431
VERSION AA579247.1 GI:2357431
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 387)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuquai,
M.D., Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Kitzman, Ph.D.
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html

Insert Length: 681 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 351.

FEATURES
Location/Qualifiers
1..387
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="915859"
/clone_lib="NCI_CGAP_P12"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: PAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected preneoplastic cells
histologically-determined to be prostatic intraepithelial
neoplasia 2 (PIN2) cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into PAMP10 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Kitzman."

BASE COUNT 125 a 54 c 67 g 141 t

ORIGIN

Query Match 97.7%; Score 374; DB 35; Length 387;
Best Local Similarity 98.4%; Pred. No. 1.3e-67;

Matches 371; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 AAAACAAACAAACAAACAAACAAATCTTCATTCGAGAAATTAATCTAGGAGCATAT 60
    |||
Db 11 AAAACAAACAAACAAACAAACAAATCTTCATTCGAGAAATTAATCTAGGAGCATAT 70
    |||
OY 61 TGTAAATATGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 120
    |||
Db 71 TGTAAATATGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 130
    |||
OY 121 TTAATAATGTCGTCGCAAAATTTTGTATTTATTTGAGAGCTTCTATCAAAAGTAATGC 180
    |||
Db 131 TTAATAATGTCGTCGCAAAATTTTGTATTTATTTGAGAGCTTCTATCAAAAGTAATGC 190
    |||
OY 181 TGGCAAGAGAGTGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 240
    |||
Db 191 TGGCAAGAGAGTGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 250
    |||
OY 241 TAAAGATTTGATTTCTCGAATGACAAATTAATTAATTAATTAATTAATTAATTAATTAAT 300
    |||
Db 251 TAAAGATTTGATTTCTCGAATGACAAATTAATTAATTAATTAATTAATTAATTAATTAAT 310
    |||
OY 301 ATAGACACACAGTCTTCTGATCTGATTAATTAATTAATTAATTAATTAATTAATTAAT 360
    |||
Db 311 ATAGACACACAGTCTTCTGATCTGATTAATTAATTAATTAATTAATTAATTAATTAAT 370
    |||
OY 361 AACATTAAGCATATATGT 377
    |||
Db 371 AACATTAAGCATATATGT 387
    |||

```

RESULT 3

```

LOCUS W56075 522 bp mRNA EST 11-OCT-1996
DEFINITION zc56b05.r1 Soares_Parathyroid_tumor_NbHPA Homo sapiens cDNA clone
IMAGE:326289 5', mRNA sequence.
W56075
ACCESSION G1357965
VERSION W56075.1 GI:1357965
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

REFERENCE

1 (bases 1 to 522)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.

The WashU-Merck EST Project
 Unpublished (1995)
 On Apr 14, 1993 this sequence version replaced gi:785211.

TITLE

JOURNAL

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LINT; contact the
 IMAGE Consortium (info@image.lim.gov) for further information.
 Insert Length: 595 Std Error: 0.00
 Seq primer: mob.REGA+ET
 High quality sequence stop: 335.
 Location/Qualifiers

FEATURES

source

1..522
 /organism="Homo sapiens"
 /db_xref="GDB:1260491"
 /db_xref="taxon:9606"
 /map="643405"
 /clone="IMAGE:326289"
 /clone_1db="Soares_Parathyroid_tumor_NbHPA"

```

/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: parathyroid gland; Vector: pTZ19D
(Pharmacia) with a modified polylinker; Site.1: Not I;
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer
[5'-
TGTTCACATCTGAAGTGGAGCGGCGGACCAATTTTCTTTTCTTTTCTTTT
T-3']], double-stranded cDNA was size selected, ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pTZ19
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Patricia Bonaldi. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases,
NIH."

```

BASE COUNT 171 a 71 c 94 g 181 t 5 others

Query Match 96.6%; Score 369.8; DB 26; Length 522;
 Best Local Similarity 97.9%; Pred. No. 8.6e-67;
 Matches 368; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

```

OY 7 AACAAACAAACAAACAAACAAATCTTCATTCGAGAAATTAATCTAGGAGCATATGTTAA 66
    |||
Db 1 AACAAACAAACAAACAAACAAATCTTCATTCGAGAAATTAATCTAGGAGCATATGTTAA 60
    |||
OY 67 TTAATGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 126
    |||
Db 61 TTAATGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 120
    |||
OY 127 TGTCTGTCGCAAAATTTTGTATTTATTTGAGAGCTTCTATCAAAAGTAATGTCGCAA 186
    |||
Db 121 TGTCTGTCGCAAAATTTTGTATTTATTTGAGAGCTTCTATCAAAAGTAATGTCGCAA 180
    |||
OY 187 AGAAGTCTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 246
    |||
Db 181 AGAAGTCTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 240
    |||
OY 247 ATTTGATTTCCGGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 306
    |||
Db 241 ATTTGATTTCCGGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 300
    |||
OY 307 CCACAGCTTCACCTTCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 366
    |||
Db 301 CCACAGCTTCACCTTCGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 360
    |||
OY 367 AAGCTAATGTTTAA 382
    |||
Db 361 AAGCTAATGTTTAA 376
    |||

```

RESULT 4

```

LOCUS A1188401/c 469 bp mRNA EST 28-OCT-1998
DEFINITION gq08e02.x1 Soares_Placenta_8to9weeks_2NBHP8to9W Homo sapiens cDNA
clone IMAGE:1723130 3', mRNA sequence.
A1188401
ACCESSION G17338610
VERSION G17338610
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

1 (bases 1 to 469)
 NCBI-GAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)

REFERENCE 1 (bases 1 to 469)
 NCBI-GAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 17, 1998 this sequence version replaced gi:1899479.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonafide, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html

Insert Length: 564 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 444.
Location/Qualifiers

1.472
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1650448"
/clone_lib="Soares.parathyroid tumor_MbHPA"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: parathyroid gland; Vector: pT73D
(Pharmacia) with a modified polylinker; Site: 1: Not I;
Site: 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer
[5',
TGTTACCAATCTGAAGTGGGAGCGCCGACCAATTTTTTTTTTTTTTTTTTTT
T-3'], double-stranded cDNA was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT73
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonafide. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases,
NIH."

BASE COUNT 175 a 80 c 55 g 162 t
ORIGIN

Query Match 87.3%; Score 334.4; DB 41; Length 472;
Best Local Similarity 97.9%; Pred. No. 1.5e-59;
Matches 332; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

45 TCTTAGGAGCTGATTTGTAATTAATGTCATTAATTTTGGGGCATTCTTA 104
|||||
DB 472 TCTTAGGAGCTGATTTGTAATTAATGTCATTAATTTTGGGGCATTCTTA 413
|||||

105 CATGCTGACAGATTAATAGTGTGTCACAAATTTTGTATTTATTTGAGACTTC 164
|||||
DB 412 CATGCTGACAGATTAATAGTGTGTCACAAATTTTGTATTTATTTGAGACTTC 353
|||||

165 TTTATCAAGATATGCTGCCAAGAGCTAAGAAATTTAGTGTGTCCTCCACACTGT 224
|||||
DB 352 TTTATCAAGATATGCTGCCAAGAGCTAAGAAATTTAGTGTGTCCTCCACACTGT 293
|||||

225 TTGAGCTGCTATTTCTAAAGATTTGATTTCTGTGAATGACATTAATTTAACTTT 284
|||||
DB 292 TTGAGCTGCTATTTCTAAAGATTTGATTTCTGTGAATGACATTAATTTAACTTT 233
|||||

285 GGTGGGGGAAAGATTAAGACACAGCTTCACTTCTATCTGTGAATTAATCTTT 344
|||||
DB 232 GGTGGGGGAAAGATTAAGACACAGCTTCACTTCTATCTGTGAATTAATCTTT 173
|||||

345 TATTGCACTGTTTGGACATTAAGCATATGTTTAAA 383
|||||

DB 172 TATTGCACTGTTTGGACATTAAGCATATGTTTAAA 134

RESULT 7
AA879150 389 bp mRNA EST 25-MAR-1998
LOCUS nw84e05.s1 NCI_CGAP_Pri12 Homo sapiens cDNA clone IMAGE:1253312,
DEFINITION mRNA sequence.
ACCESSION AA879150
MIM 92968115
VERSION AA879150.1 GI:2988115
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 389)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2153434.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D.,
Rodrigo F. Chuang, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 381.
Location/Qualifiers

1.389
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="12"
/clone="IMAGE:1253312"
/clone_lib="NCI_CGAP_Pri12"
/sex="male"
/tissue_type="metastatic prostate bone lesion"
/lab_host="DH10B"
/note="Vector: PAMP10; mRNA made from metastatic prostate
lesion of the bone, cDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 600 bp. Library made by D. Kitzman,
NIH."

BASE COUNT 121 a 51 c 75 g 142 t
ORIGIN

Query Match 85.5%; Score 327.4; DB 39; Length 389;
Best Local Similarity 95.5%; Pred. No. 4.2e-58;
Matches 342; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

26 TCTTCATTTAGAAAATTAATCTTAGGAGCTGATTTGTAATTAATGTCATTAATTT 85
|||||
DB 3 TCGACATTTAGAAAATTAATCTTAGGAGCTGATTTGTAATTAATGTCATTAATTT 62
|||||

86 RTTGTGGGCACTTCTTCAATGCTTGTGACAAAGATTAATAGTGTGTCGCAAAATTTTG 145
|||||
DB 63 ATT-TGGGCACTTCTTCAATGCTTGTGACAAAGATTAATAGTGTGTCGCAAAATTTTG 121
|||||

146 TATTATTATTGAGACTTCTTATCAAAAGTAATGCTGCCAAGAGAGCTAAGAAATTAG 205
|||||
DB 122 TATTATTATTGAGACTTCTTATCAAAAGTAATGCTGCCAAGAGAGCTAAGAAATTAG 181
|||||

206 TAGTGTCCCMTCACCTGTTTGGAGTGTGCTATTTCTAAAGATTTTGAATTTCTCGAATG 265
|||||

```

|||||
Db 182 TAGTCTCCACACTGTTGTTGAGTGCATTCATAAAGTTTGATTGTTCTGGAAG 241
Oy 266 ACAATTATATTTAACTTTGGTGGGGAANAAGTATAGACCAAGCTTCACCTTGA 325
Db 242 ACAATTATATTTAACTTTGGTGGGGAANAAGTATAGACCAAGCTTCACCTTGA 301
Oy 326 TACTGTAAATATATCTTTATTTGCACTGTTTGACATTAAGCATATGTTTAAA 383
Db 302 TACTGTAAATATATCTTTATTTGCACTGTTTGACATTAAGCATATGTTTAAA 359

RESULT 8
A1160743 442 bp mRNA EST 26-OCT-1998
LOCUS qB49d03.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:1703429 3',
DEFINITION mRNA sequence.
ACCESSION A1160743
VERSION 93694123
KEYWORDS A1160743.1 GI:3694123
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 442)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(JCAP/BTGP), Tumor Gene Index
Unpublished (1998)
JOURNAL On Jan 19, 1998 this sequence version replaced gi:2286564.
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonald, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 2616 Std Error: 0.00
Seq primer: -40m13 fwd. Et from Amersham
High quality sequence stop: 433.
Location/Qualifiers
1. 442
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1703429"
/clone_lib="NCI_CGAP_Brn23"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTGAGTGGAGCGCGCCATATCTTTTATTTTATTTTATTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonald."

BASE COUNT 172 a 76 c 53 g 141 t
ORIGIN
Query Match 83.7%; Score 320.6; DB 43; Length 442;
Best Local Similarity 97.0%; Pred. No. 1e-56;
Matched 320; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

```

```

Oy 54 CTGATATTGTAATTATNGTCATTTAATWRTTKTGGGGCATTTCTTACATTGCTT 113
Db 442 CTGATATTGTAATTATNGTCATTTAATWRTTKTGGGGCATTTCTTACATTGCTT 383
Oy 114 GACAAGATTAATAATGTCGTGCCAAAATTTTGAATTTATTTGAGAGACTTTATCAAA 173
Db 382 GACAAGATTAATAATGTCGTGCCAAAATTTTGAATTTATTTGAGAGACTTTATCAAA 323
Oy 174 GTAATGCTGCCAAGAGAGTATAGAAATTAAGTAGTGTCCCMTCACCTGTTGGAGTGT 233
Db 322 GTAATGCTGCCAAGAGAGTATAGAAATTAAGTAGTGTCCCMTCACCTGTTGGAGTGT 263
Oy 234 GGTATTCATAAAGATTGATTCTGGAATGACATTAATTTAATCTTGTGGGGGCA 293
Db 262 GGTATTCATAAAGATTGATTCTGGAATGACATTAATTTAATCTTGTGGGGGCA 203
Oy 294 AANAATTATAGACCAAGCTTCTGACTTCTGATTAATTAATCTTTATTTGCACT 353
Db 202 AANAATTATAGACCAAGCTTCTGACTTCTGATTAATTAATCTTTATTTGCACT 143
Oy 354 TGTTTTGACATTAAGCTATATGTTTAAA 383
Db 142 TGTTTTGACATTAAGCTATATGTTTAAA 113

RESULT 9
A1522136 444 bp mRNA EST 13-APR-1999
LOCUS t178f03.x1 NCI_CGAP_K1d11 Homo sapiens cDNA clone IMAGE:2138141 3',
DEFINITION mRNA sequence.
ACCESSION A1522136
VERSION 94436271
KEYWORDS A1522136.1 GI:4436271
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 444)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL On Mar 10, 1998 this sequence version replaced gi:2948750.
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 539 Std Error: 0.00
Seq primer: -40UP from GIBCO.
Location/Qualifiers
1. 444
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2138141"
/clone_lib="NCI_CGAP_K1d11"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_K1d3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs

```

Page 7

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523</
--	---	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-------

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523</
--	---	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-------

vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682652-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 193 a 79 c 57 g 167 t
ORIGIN

Query Match 79.2%; Score 303.4; DB 41; Length 496;
Best Local Similarity 97.5%; Pred. No. 3.2e-53;
Matches 312; Conservative 5; Mismatches 2; Indels 1; Gaps 1;

QY 64 TAATTGTCATTTAAATTTTGGGCGCTTCTTACCTTCTTGGACAATTA 123
|||||
DB 496 TAATTATGTCATTTAAATTTTGGGCGCTTCTTACCTTCTTGGACA-ATTA 438
|||||
QY 124 AATGCTGTCGCAAAATTTTGTATTTTGGAGCTTCTTATCAAAAGTATGCTGC 183
|||||
DB 437 AATGCTGTCGCAAAATTTTGTATTTTGGAGCTTCTTATCAAAAGTATGCTGC 378
|||||
QY 184 CAAGAGAGCTTAAGCAATTAAGTATGCTTCCCTCACTTGTGGAGTGTCTATCTAA 243
|||||
DB 377 CAAGAGAGCTTAAGCAATTAAGTATGCTTCCCTCACTTGTGGAGTGTCTATCTAA 318
|||||
QY 244 AAGATTTTATTTTCCGTAATGCAATTTATTTTACCTTGTGGGGAAGTATTA 303
|||||
DB 317 AAGATTTTATTTTCCGTAATGCAATTTATTTTACCTTGTGGGGAAGTATTA 258
|||||
QY 304 GGACCACTCTTCACTTGTGATCTTGAATTAATCTTTTATGCACTGTTTGAAC 363
|||||
DB 257 GGACCACTCTTCACTTGTGATCTTGAATTAATCTTTTATGCACTGTTTGAAC 198
|||||
QY 364 ATTAAGCTATATGTTTAA 383
|||||
DB 197 ATTAAGCTATATGTTTAA 178
|||||

RESULT 12
AI051146 473 bp mRNA EST 24-SEP-1998
LOCUS 0Y49D04.X1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:1669159 3',
DEFINITION mRNA sequence.
ACCESSION AI051146
NID 93306680
VERSION AI051146.1 GI:3306680
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 473)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/RTGAP), Tumor Gene Index
UNPUBLISHED (1998)

JOURNAL
COMMENT On Jan 17, 1998 this sequence version replaced gi:2044831.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-rl0.llnl.gov/dbp/image/image.html

Insert Length: 2319 Std Error: 0.00
Seq primer: -40m13 fwd. ER from Amersham
High quality sequence stop: 438.
Location/Qualifiers
1. 473
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1669159"
/clone_lib="NCI_CGAP_Brn23"
/tissue_type="g1oblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTCACATCTGAGTGGAGCGCGCGCATATCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."

BASE COUNT 181 a 77 c 56 g 159 t
ORIGIN

Query Match 76.1%; Score 291.6; DB 41; Length 473;
Best Local Similarity 98.6%; Pred. No. 8.4e-51;
Matches 291; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 89 KTGGGGCAATTCCTTACATTTGCTTGACAGATTTAAATGCTGCGCAAAATTTGTAT 148
:|||||
DB 473 TGGGGCAATTCCTTACATTTGCTTGACAGATTTAAATGCTGCGCAAAATTTGTAT 414
:|||||
QY 149 TTTATTTGAGACTTCTTATCAAAAGTATGCTGCCAAAGAGTCTAAGCAATTAAG 208
|||||
DB 413 TTTATTTGAGACTTCTTATCAAAAGTATGCTGCCAAAGAGTCTAAGCAATTAAG 354
|||||
QY 209 TGTCCCACTCTTGTGGAGTGTCTATCTAAAGATTTGATTCCTGGAATGACA 268
|||||
DB 353 TGTCCCACTCTTGTGGAGTGTCTATCTAAAGATTTGATTCCTGGAATGACA 294
|||||
QY 269 ATTATATTTTAACTTTGGTGGGGAAGTATAGGACACAGCTCTCACTTGTATAC 328
|||||
DB 293 ATTATATTTTAACTTTGGTGGGGAAGTATAGGACACAGCTCTCACTTGTATAC 234
|||||
QY 329 TTGTAAATTAATCTTTATGCACTTGTGACCATTAAGCTATATGTTTAA 383
|||||
DB 233 TTGTAAATTAATCTTTATGCACTTGTGACCATTAAGCTATATGTTTAA 179
|||||

RESULT 13
AA622892 407 bp mRNA EST 21-OCT-1997
LOCUS np58C12.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1130518 3',
DEFINITION similar to contains Alu repetitive element, mRNA sequence.
ACCESSION AA622892
NID 92526768
VERSION AA622892.1 GI:2526768
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 407)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
UNPUBLISHED (1997)

JOURNAL
COMMENT On May 18, 1995 this sequence version replaced gi:810969.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:
www-bio.lnl.gov/db/rlp/image/image.html

Insert Length: 948 Std Error: 0.00
 Seq. primer: -40m13 fwd. RT from Amersham
 High quality sequence stop: 385.
 Location/Qualifiers

FEATURES

source

1. 407
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1130518"
 /clone_lib="NCI-CGAP_Br2"
 /sex="female, pooled"
 /tissue_type="breast"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled bulk breast tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT733 vector. This library is the normalized version of NCI-CGAP_Br1.1. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 157 a 72 c 50 g 128 t
 ORIGIN

Query Match 75.2%; Score 288; DB 36; Length 407;
 Best Local Similarity 99.0%; Pred. No. 4.7e-50;
 Matches 288; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

93 GGCATTCCTTACATGCTTGACCAAGATTAATGCTGCGCAAAATTTGATTTA 152
 |||
 DB 407 GGCATTCCTTACATGCTTGACCAAGATTAATGCTGCGCAAAATTTGATTTA 348
 |||
 QY 153 TTGGAGACTCTTCAAAAAGTAATGCTGCCAAGAGAGCTTAAGAAATTAAGTGT 212
 |||
 DB 347 TTGGAGACTCTTCAAAAAGTAATGCTGCCAAGAGAGCTTAAGAAATTAAGTGT 288
 |||
 QY 213 CCCATCCTGTTGGAGTGTCTATCTTAAGATTTGATTTCTCGAATGCAATTA 272
 |||
 DB 287 CCCATCCTGTTGGAGTGTCTATCTTAAGATTTGATTTCTCGAATGCAATTA 228
 |||
 QY 273 TATTTAACTTGGTGGGGAANAAGTATAGGACCACTCTTCACTTGTGATCTGT 332
 |||
 DB 227 TATTTAACTTGGTGGGGAANAAGTATAGGACCACTCTTCACTTGTGATCTGT 168
 |||
 QY 333 AAATTAATCTTTATGCACTGTTTGGACCAATTAAGCTATATGTTAA 383
 |||
 DB 167 AAATTAATCTTTATGCACTGTTTGGACCAATTAAGCTATATGTTAA 117
 |||

RESULT 14

LOCUS C14656 344 bp mRNA EST 30-SEP-1996
 DEFINITION C14656 Clontech human aorta polyA+ mRNA (#6572) Homo sapiens CDNA
 clone GEN-078E02 5', mRNA sequence.
 ACCESSION C14656
 NID 91569363
 VERSION C14656.1 GI:1569363
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 344)

AUTHORS Fujiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M., Okuno, S., Ozaki, R., Shimizu, F., Shimada, Y., Shinomura, H., Takachi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y., Maekawa, H., Shin, S. and Nakamura, Y.
 TITLE
 JOURNAL
 COMMENT
 On Sep 12, 1996 this sequence version replaced gi:1407498.

Contact: Tsutomu Fujiwara
 Otsuka GEN Research Institute
 Otsuka Pharmaceutical Co., Ltd
 463-10 Kagasuno Kawachi-cho, Tokushima, Tokushima, 771-01 Japan
 Tel: 0886-65-2888
 Fax: 0886-37-1035.

FEATURES

source

1. 344
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="GEN-078E02"
 /clone_lib="Clontech human aorta polyA+ mRNA (#6572)"

BASE COUNT 138 a 44 c 50 g 108 t 4 others
 ORIGIN

Query Match 73.2%; Score 280.4; DB 28; Length 344;
 Best Local Similarity 96.5%; Pred. No. 1.7e-48;
 Matches 276; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAACAACAAACAAAAACAAATCTTCATTCAGAAAAATTAATCTTGAAGAGATAT 60
 |||
 DB 59 AAAACAACAAACAAAAACAAATCTTCATTCAGAAAAATTAATCTTGAAGAGATAT 118
 |||
 QY 61 TGGTATATAGTCATTTAATATATTTTGGGCACTTCTTACATTTCTTGACAA 120
 |||
 DB 119 TGGTATATAGTCATTTAATATATTTTGGGCACTTCTTACATTTCTTGACAA 178
 |||
 QY 121 TTAATATGCTGCGCAAAATTTTGAATTTAATTTGAGACCTCTATCAAAAGTAATGC 180
 |||
 DB 179 TTAATATGCTGCGCAAAATTTTGAATTTAATTTGAGACCTCTATCAAAAGTAATGC 238
 |||
 QY 181 TGGCAAGAGAGCTTAAGAAATTAAGTGTGCCATCTGTTGAGTGTCTATTC 240
 |||
 DB 239 TGGCAAGAGAGCTTAAGAAATTAAGTGTGCCATCTGTTGAGTGTCTATTC 298
 |||
 QY 241 TAAATATTTGATTTCTCGAATGCAATTAATTTTAACTTTGG 286
 |||
 DB 299 TAAATATTTGATTTCTCGAATGCAATTAATTTTAACTTTGG 344
 |||

RESULT 15

LOCUS A1420653 463 bp mRNA EST 28-MAR-1999
 DEFINITION A1420653 t113b04.x1 NCI-CGAP_Br23 Homo sapiens CDNA clone IMAGE:2096047 3', mRNA sequence.
 ACCESSION A1420653
 NID 94266584
 VERSION A1420653.1 GI:4266584
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 463)
 NCI/NCI/NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BRGAP), Tumor Gene Index

Unpublished (1998)
 On Mar 20, 1998 this sequence version replaced gi:2980399.

CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov


```

source
1..197441
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="bac R-1012A1 from RPOI-11 library"
BASE COUNT      57409 a 42215 c 42060 g 55757 t
ORIGIN

Query Match      98.8%; Score 378.4; DB 34; Length 197441;
Best Local Similarity 98.2%; Pred. No. 5,3e-61;
Matches 376; Conservative 5; Mismatches 2; Indels 0; Gaps 0

QY      1  AAAACACACAAACAAAAAACAAATCTTCATTCAGAAAAATATCTTGAAGGACTATAT 60
Db 50969  AAAACAAACAAACAAAAACAAAAACAAATCTTCATTCAGAAAAATATCTTGAAGGACTATAT 50910

QY      61  TGGTATATATGTCATTTAATATTTTGTGGGGCACTTCCTACATGCTCGAAGAAG 120
Db 50909  TGGTATATATGTCATTTAATATATTTTGGGGCACTTCCTACATGCTCGAAGAAG 50850

QY      121  TTAATAATGCTGTGCCAAATATTTTGATTTTATTTGAGAGCTCTATCAAAAGTATGC 180
Db 50849  TTAATAATGCTGTGCCAAATATTTTGATTTTATTTGAGAGCTCTATCAAAAGTATGC 50790

QY      181  TGCCAAAGAGAGCTAAGGAATTAGTAGTGTCCCTCACTGTTTGGAGTGTGCTATTC 240
Db 50789  TGCCAAAGAGAGCTAAGGAATTAGTAGTGTCCCTCACTGTTTGGAGTGTGCTATTC 50730

QY      241  TAAAGATTTTGAATTCCTCGAATGACAAATATATTTTAACTTGTGGGGGAANAAGTT 300
Db 50729  TAAAGATTTTGAATTCCTCGAATGACAAATATATTTTAACTTGTGGGGGAANAAGTT 50670

QY      301  ATAGGACCAACAGCTCTTCACCTTCTGATACCTGTAAATATCTTTATTCACCTTGTTTG 360
Db 50669  ATAGGACCAACAGCTCTTCACCTTCTGATACCTGTAAATATCTTTATTCACCTTGTTTG 50610

QY      361  ACCATTAGCTATATGTTTAAAAA 383
Db 50609  ACCATTAGCTATATGTTTAAAGAA 50587

```

	RESULT	2	
	HSAC002076/c		
LOCUS	HSAC002076	149764 bp	DNA
DEFINITION	Human BAC clone GS345D13 from Tq31-q32,	PRI	12-MAY-1997
ACCESSION	AC002076		complete sequence.
NID	G2078461		
VERSION	AC002076.1	GI:2078461	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Homidae; Homo.		
TITLE	Magnifl.		
JOURNAL REFERENCE	The sequence of H. sapiens BAC clone GS345D13 unpublished (1997)		
AUTHORS	2 (bases 1 to 149764)		
TITLE	Waterston,R.		
JOURNAL COMMENT	Direct Submission Submitted (12-MAY-1997) SUBMITTED BY:		
	Genome Sequencing Center Department of Genetics Washington University St. Louis MO 63108, USA http://genome.wustl.edu/gsc mailto:sapiens@watson.wustl.edu		

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap

between neighboring data submissions

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION: The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHRG1 Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhrgr1.nih.gov/DIR/GRB/CHK/> or send an E-mail to green@nhrgr1.nih.gov

Mapping information for this clone was also provided by Dr. John D. McPherson, Department of Genetics/Genome Sequencing Center, Washington University School of Medicine.

SOURCE INFORMATION:
This clone is from the first BAC library from Genome Systems, Inc.
(<http://www.genomesystems.com>).
Cell line: Jumbob12c04

Cell line: lymphoblastoid
Haplotypes: two
VECTOR: pBelO-BAC
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is H_G5274A07, 200 bp overlap. Actual start of this clone is at base position 1 of H_G5345D13. The actual end is at 19827 of H_G5274A07. This cosmid lies in an unanchored cluster of unknown orientation.

This clone contains STS SWSS4055 (NID:g1916534), SWSS2840 (NID:g113614), and SWSS1647 (NID:g113169).

FEATURES

Source

1. .149764

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"
```

```
/chromosome="7"  
/clone="GS345D13"
```

```
/clone_11b="GSBAC1"
/map="7q31-q32"
```

```
3138. .4124
/rpt_family="L1"
```

```
4162. .4237
/rpt_family="L1"
```

```

/1P_family= LI
complement(4345. .4638)
/rpt_family="ALT"

```

```
/ipl_railly=ALU
4769. :5162
/note="match to
```

```

/notice=match to
zk39d04.s1"
comp]ement(5)19

```

```
complement(3119.  
/note="match to  
zk39d04 r1"
```

zk39d04.tl
6029. .6148

```

/irpt_family="L1"
6787. .7179

```

```
/note="match to
z147c03.s1"
```

```
6788. .7175
/note="match to
```

z122b10.s1"
6788. .7201

```
/note="match to  
yy08a02.s1"
```

```
6802. :7102
/note="match to
```

```

/note=match to
6909.7179

```

complement(7074.

[illegible]

contaminated with foreign sequence from E.coli, yeast, vector, phage etc.

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source
Location/Qualifiers
1..207139
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="II"
/clone="Y17G7"

BASE COUNT 68582 a 34583 c 35567 g 67606 t 801 others
ORIGIN

Query Match 11.8%; Score 45.2; DB 34; Length 207139;
Best Local Similarity 49.5%; Pred. No. 1.8;
Matches 140; Conservative 1; Mismatches 140; Indels 2; Gaps 1;

QY 97 TTTCCTTACATGTCCTGACAGATTAATAATGCTGTCGCAAAATTTGATTTATTTG 156
DB 76011 TTTCCTTACATGTCCTGACAGATTAATAATGCTGTCGCAAAATTTGATTTATTTG 156
QY 157 GAGACTTCTTATCAAAAGTAATGCTGCCAAGAGCTTAAGATTAAGTCTCCM 216
DB 76071 GAGGCTTCGATACCAATATCAATGCAAAAAATTCGATTTTATTA 76130
QY 217 TCACCTGTTGAGCTGCTATTCCTAAGATTTGATTTCTCGAATGCAATTAAT 276
DB 76131 TTAATTTTAAATCCAAAAATCCATGCTTTTAAATTC- AACGCAATTTTGA 76188
QY 277 TTAACCTTGTGGGGAANAGTTATAGCCAGCTTCTGATCTGATCTGAAT 336
DB 76189 CTGATTACCGTCCGAGGAGTGTGATTTAGCAATTTATTTATTTATTTT 76248
QY 337 TAATCTTTATGCACTGTTTGGACCAATTAAGCTATATGTT 379
DB 76249 AAAATTTTCAACCAATTTCTGATTTCTGCTTTTGT 76291

RESULT 4
CEY17G7B 143092 bp DNA INV 30-NOV-1998
LOCUS Caenorhabditis elegans cosmid Y17G7B, complete sequence.
DEFINITION AL023828
ACCESSION 93217816
NID AL023828.1 GI:3217816
VERSION HTG.
KEYWORDS
SOURCE Caenorhabditis elegans.
ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditiida; Rhabditiina; Rhabditiidae; Rhabditiidae; Pelodermidae; Caenorhabditis.

REFERENCE 1 (bases 1 to 143092)
AUTHORS Smye, R.
TITLE Direct Submission
JOURNAL Submitted (09-JUN-1998) Louis, MO 63110, USA. E-mail:
jesesanger.ac.uk or rwenematode.wustl.edu

REFERENCE 2 (bases 1 to 143092)
AUTHORS Wilson, R., Almscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
Lattelle, P., Lightning, J., Lloyd, C., McQuarrie, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Rikken, L., Roopra, A.,
Saunders, D., Showken, R., Smaildon, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Spratt, J., and Woldman, P.

TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL Nature 368 (6466), 32-38 (1994)

MEDLINE
COMMENT

94150718
Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see: "
http://webc.sanger.ac.uk/Cq1-
bin/displayrb-wormaccetclass-sequence-objct-y17g7b
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone Y17G7B.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true right end of clone W03C9 is at 100 in this sequence. The
true right end of clone Y17G7 is at 143092 in this sequence. The
start of this sequence (1..100) overlaps with the end of sequence
Z66516.
The end of this sequence overlaps with the start of sequence
AL022279.

FEATURES
source

Location/Qualifiers
1..143092
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="II"
/clone="Y17G7B.1"
complement(3569..4874)
/gene="Y17G7B.1"
/gene="Y17G7B.1"
/note="similar to Acyl CoA binding protein"
/codon_start=1
/protein_id="CA19448.1"
/db_xref="PID:e1333259"
/db_xref="PID:g3947596"
/db_xref="GI:3947596"
translation="MPISLEKMTGGLFEFAEEMRLKSEPTDRERKLYALYKOA
LHGDIPIEDVYVPAAGDEVGKTKAAKMSQGANSEKCRADYVAIAEMIKRYGRNIV
KCKMSEWASYDY"
11683..19658
/gene="Y17G7B.2b"
11683..19658
/gene="Y17G7B.2c"
11683..19658
/gene="Y17G7B.2a"
join(11683..11753,11836..11928,11980..12304,13242..13462,
14218..14331,15865..16127,17187..17299,18228..18428,
18624..18714,19438..19658)
/gene="Y17G7B.2c"
/note="Y17G7B.2a"
/codon_start=1
/protein_id="CA19449.1"
/db_xref="PID:e1333260"
/db_xref="PID:g3947597"
/db_xref="GI:3947597"
translation="WRSSKGRGQAAKPTATTCVCDGREGSEVYVSTCLKMF
HGCKLKEHEINSGPWPICITFTTCQCRPTADWAKKADLVOMCVTATLSAER
LKAGCKSAEHPDFTYLSLKDIEIVPMENMYMLAIAKCKEMHOLAPTLKERN
GAFVRHDEADDTFAAEKMLSLGDLHEAVKLIGKRLERERERHLELPIDEPKTR
ISKRNHAEADDTFAAEKMLSLGDLHEAVKLIGKRLERERERHLELPIDEPKTR
AMNONSAAVYIPSEHYRELINPUNVANSNDRAOLSTKSGYREGSMAARASHV
SKGTYFEVNDPDDSHIRIKGSOSASIALOACVGYKREYGRKRGTFHFAKCK
KYHFGFKQSDVGLGLHLIPVKKLIQIPANIPSEKYLPSVSKGNLSIFKANYFEVQ
EESADIAITLVEMPGSYIEFPHNKSCKGAENYAGAYPSISIFSATPTMMLGPK
FRNLPKATGTHARADEQHQBTLSMDLIVTSKEVNDHPVPRVAREDDDVKDIKEI"
KOEI"

JOURNAL Unpublished (1999)
 REFERENCE 2 (bases 1 to 104308)
 AUTHORS Ecker, J.R.
 TITLE Direct Submission
 JOURNAL Submitted (18-DEC-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA

REFERENCE 3 (bases 1 to 104308)
 AUTHORS Ecker, J.R.
 TITLE Direct Submission
 JOURNAL Submitted (11-APR-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
 On Apr 11, 1999 this sequence version replaced gi:4389503.

COMMENT Location/Qualifiers
 FEATURES
 source 1..104308
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="1"

BASE COUNT 33546 a 18958 c 18485 g 33319 t
 ORIGIN

Query Match 11.7%; Score 45; DB 8; Length 104308;
 Best Local Similarity 48.5%; Pred. No. 2.4;
 Matches 114; Conservative 3; Mismatches 118; Indels 0; Gaps 0;

OY 19 AATCAATCTTCAATTCAGAAAAATATCTTAGGAGCTATATGTAATTAATGTCATTT 78
 DB 56700 AATTAATCTTCAATTCAGAAAAATATCTTAGGAGCTATATGTAATTAATGTCATTT 56641
 OY 79 TAAATRTTGTGGGAGCATTTCTTACATTTGCTGACAAAGATTAATTAATGTCGCCAA 138
 DB 56640 TTGTAATTAATTAAGTAATCTTTTAATTAATTAATTAATTAATTAATTAATTAATTA 56581
 OY 139 AATTTGATTTTATTTGAGAGCTCTTATCAAAAGTAATGTCGCCAAGAGAGCTAAG 198
 DB 56580 GACTCTCTACATTAACATACACATAGTTAGTAATAAATTAATTAATTAATTAATTAATTA 56521
 OY 199 GAATTAGTAGTGTCTCCATCTGTTTGGAGTGTCTTATCAAAAGATTTTGA 253
 DB 56520 TATTTAAATAATTAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 56466

RESULT 6
 PPMALIP2/c PPMALIP2 82044 bp DNA HTG 09-APR-1999
 LOCUS Plasmodium falciparum chromosome 1 strain 3D7, WORKING DRAFT
 DEFINITION
 SEQUENCE, in unordered pieces.
 ACCESSION AL031745
 NID 94493855
 VERSION AL031745.3 GI:4493855
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE malarial parasite P. falciparum.
 ORGANISM Plasmodium falciparum
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 82044)
 Bowman, S., Churcher, C., Harris, D., Lawson, D., Quail, M. and Barrett, B.
 TITLE Direct Submission
 JOURNAL Submitted (09-APR-1999) P. falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
 COMMENT On Mar 24, 1999 this sequence version replaced gi:4455716.
 For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc.
 Order of segments is not known; 800 n's separate segments.

* NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES
 source Location/Qualifiers
 1..82044
 /organism="Plasmodium falciparum"
 /strain="3D7"
 /db_xref="taxon:5833"
 /chromosome="1"

BASE COUNT 27301 a 7762 c 6747 g 29034 t 11200 others
 ORIGIN

Query Match 11.6%; Score 44.6; DB 34; Length 82044;
 Best Local Similarity 46.7%; Pred. No. 5;
 Matches 134; Conservative 2; Mismatches 151; Indels 0; Gaps 0;

OY 1 AATCAACAAACAAACAAACAAATTCCTTCATTCGAAATTAATCTTAGGAGCTATAT 60
 DB 46959 AAAAAAAAAAAAAAAAAAAAAAAAAATATATATATATATATATATATATATATATAT 46900
 OY 61 TGGTAATATAGTCAATTTAAATWRTTKTGCGGCATTTCTTACATTTGCTGCAAGA 120
 DB 46899 ATTACAT 46840
 OY 121 TTAATATGCTGTGCCAAATTTTGTATTTATTTGGAGACTCTTANCAAAAGTAATGC 180
 DB 46839 TTTTATGATTTTATTTACAAATATATATATATATATATATATATATATATATATATG 46780
 OY 181 TGCCAAAGAGATCTAAGAAATTAAGTATAGTGTCCCTGACTGTTTGGAGTGTATTC 240
 DB 46779 TTATATGCTGTAGAGACTATATAAGTATAGTATATAACAAAGCATGATGAAATAT 46720
 OY 241 TAAAGATTTTGATTTCTGGAATGACATTAATATTTTAACCTTGGT 287
 DB 46719 TATTAAGATGTTACACATGTTAATATATATATGTTTATATATATG 46673

RESULT 7
 IP006960
 LOCUS IP006960 441 bp DNA INV 18-FEB-1995
 DEFINITION Ichneumon promisorius mitochondrial 16S rRNA gene, partial
 SEQUENCE.
 ACCESSION U06960
 NID 9463865
 VERSION U06960.1 GI:463865
 KEYWORDS
 SOURCE Ichneumon promisorius.
 ORGANISM Ichneumon promisorius
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Plekoyota; Neoptera; Endopterygota; Hymenoptera; Apocrita;
 Ichneumonidae; Ichneumonidae; Ichneumoninae; Ichneumon.
 1 (bases 1 to 441)
 Dowton, M. and Austin, A.D.
 TITLE Molecular phylogeny of the insect order Hymenoptera: apocritan
 relationships
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (21), 9911-9915 (1994)
 MEDLINE 95024071
 REFERENCE 2 (bases 1 to 441)
 AUTHORS Dowton, M.
 TITLE Direct Submission
 JOURNAL Submitted (21-FEB-1994) Mark Dowton, Department of Crop Protection,
 Waite Campus, University of Adelaide, Adelaide, South Australia
 5064, Australia
 FEATURES
 source Location/Qualifiers
 1..441
 /organism="Ichneumon promisorius"
 /mitochondrion
 /db_xref="taxon:32393"
 /dev_stage="adult"
 <1..>441
 /note="large subunit ribosomal RNA"

QY 220 CTTGTTGGAGTGTCTATTCTAAAGATTTCCTCGATGACATTTATTTTA 279
 1 111 1 111 1 111 1 111 1 111 1 111 1 111 1
 Db 66099 AATATTTTACTTTCTCTGAGCATTTTCAGTTAGCCTTGATGACGTATATCATGT 66040
 QY 280 ACTTTG 285
 111111
 Db 66039 ACTTTG 66034
 RESULT 10
 AC006259
 LOCUS AC006259 110680 bp DNA PLN 28-DEC-1998
 DEFINITION Arabidopsis thaliana BAC F2J16 from chromosome V, containing KNAT3
 and mapping near 60.5 cM, complete sequence.
 ACCESSION AC006259
 NID 94063730
 VERSION AC006259.1 GI:4063730
 KEYWORDS HTG.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsis.
 REFERENCE 1 (bases 1 to 110680)
 Spiegel, L.A., de la Bastide, M., Habermann, K., Shah, R.,
 Rodriguez, M., Shekher, M., Swaby, I., Matero, A., Preston, R.,
 Vill, M.D., Schutz, K., Nascimento, L., Huang, E.N., Dedhia, N.,
 Parnell, L.D. and McCombie, W.R.
 Arabidopsis thaliana BAC F2J16 from chromosome V, near 60.5 cM
 Unpublished
 TITLE 2 (bases 1 to 110680)
 Spiegel, L.A., de la Bastide, M., Habermann, K., Shah, R.,
 Rodriguez, M., Shekher, M., Swaby, I., Matero, A., Preston, R.,
 Vill, M.D., Schutz, K., Nascimento, L., Huang, E.N., Dedhia, N.,
 Parnell, L.D. and McCombie, W.R.
 REFERENCE JOURNAL
 AUTHORS Submitted (28-DEC-1998) Lita Annenberg Hazen Genome Sequencing
 Center, Cold Spring Harbor Laboratory, PO Box 100, Cold Spring
 Harbor, NY 11724, USA
 COMMENT BAC F2J16 is assigned to YAC C1C2B3 and maps to near 60.5 cM on the
 Lister & Dean RI map. Position 1 of F2J16, the T7 end, is oriented
 toward the telomere and position 110680, the SP6 end, is oriented
 toward the centromere. For more information on the mapping,
 sequencing and annotation of F2J16, please see
 http://www.cshl.org/Arabidopsis/F2J16-titlepage.html. A graphic view of
 our annotation is also available at this url. Gene models are built
 with exons predicted by GENSCAN
 (http://ccr-081.mit.edu/GENSCAN.html), MZEF
 (http://www.cshl.org/genefinder) and GRATL
 (http://compbio.ornl.gov/tools/index.shtml) and with splice sites
 predicted by NetPlantGene
 (http://www.cbs.dtu.dk/netgene/cbsnetgene.html). Alternate exons
 not used in building the gene models are presented on the web pages
 associated with F2J16. Genes are numbered according to the scheme
 BAC.gene_number. Typically, these numbers progress from 1 upwards
 as one moves from position 1 of the BAC. Protein sequences coded
 by the genes are assigned to a functional category with the aid of
 similarity searches and comparison to the Prosite
 (http://expasy.hcuge.ch/prosite/prosite.html) and Pfam
 (http://pfam.wustl.edu/) libraries. A description of these
 categories can be found at
 http://manijac.mips.biochem.mpg.de/Arabidopsis/. Genomic repeats are
 typically located by TBLASTX analysis and an attempt is made to
 classify the function of each repeat as either transposon, putative
 microsatellite, LINE, direct repeat, centromeric repeat, etc.
 If you have any questions or confirmatory or contradictory evidence
 concerning the annotation of F2J16, please direct email to Larry
 Parnell at parnell@cshl.org.
 FEATURES
 source 1. 110680
 /organism="Arabidopsis thaliana"

gene
 /cultiyar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="V"
 /map="near 60.5 cM"
 <1..717
 /gene="F2J16.1"
 /note="encodes putative P450; entire gene encoded on
 T11H3, Genbank accession number AC005964; genomic copy of
 EST Z33677; genomic copy of EST Z33963; gene model last
 edited on 16 Dec 98"
 /evidence="not_experimental"
 1..24515
 /note="overlap with T11H3, Genbank accession number
 AC005964, from position 67404 to 91916"
 /note="region: overlap with T11H3"
 join(<1..11,103..717)
 /gene="F2J16.1"
 /note="functional catalog ID=98"
 /codon_start=3
 /evidence="not_experimental"
 /product="putative P450"
 /protein_id="AAC98443.1"
 /db_xref="PID:94063733"
 /db_xref="GI:4063733"
 /translation="ILFVILNAGIDTSAQVWTWYTLISNPRVAKKAQAEVREVIN
 KDIIEEDIERLEFLYKMWIKETFRVPLIPVLLIPRAKSKVYKIGYDIPKKTWVHNT
 WAHRRNPRVWDPPEALFPERFMDQIDIKGLNFELPPGSGRRCPGIGMGMAVHLT
 LILNLYRDMWLPESMEVEDLESTGLVCPKRPVLOLIPVLITQWS"
 complement(1310..1353)
 /note="function-putative_microsatellite"
 /rpl_type=tandem
 /rpl_unit=AT
 4808..6391
 /gene="F2J16.2"
 /note="encodes putative P450; gene model last edited on 16
 Dec 98"
 /evidence="not_experimental"
 join(4808..5683,5777..6391)
 /gene="F2J16.2"
 /note="functional catalog ID=98"
 /codon_start=1
 /evidence="not_experimental"
 /product="putative P450"
 /protein_id="AAC98444.1"
 /db_xref="PID:94063734"
 /db_xref="GI:4063734"
 /translation="MSLWYIIIVAFVFFSSMIYVIRKTKNLPPGPPRLPTIGNLHQ
 LGSFPHSMFRLSETYGPIMSLKRGSVSTVVASPEVYKTFEVECCSRPMWTP
 ARVYNYKIDCFSPYSKRYRPMVRYVETLTARVQSFQRTREVEVAALVDFIKQAA
 SLERPVNLKMLKLSGVICRVAFGINLOSKELENTYEYVIOCTVELVGSFAADYF
 PVARIIIDRTIGLSKCEKFKRADAFPDOSIKHLDEILIKDIIIDLLKMEGETT
 LGSEQLRDHRTKGIANTLNGIDITSQVMTWVYTLISNPRVAKKAQAEVREYIKR
 DDILEEDIERLQYLVKIKETFRINPLIPVLLIPRAKSKVYKIGYDIPKKTWVHNT
 AINLRYRDMWLPESMEVEDLESTGLVCPKRPVLOLIPVLITQWS"
 23757..24702
 /gene="F2J16.6"
 /note="mRNA sequence based on a comparison to cDNA L39646"
 23757..24702
 /gene="F2J16.6"
 /note="encodes ZFP3 zinc finger protein; genomic copy of
 cDNA L39646"
 /evidence="experimental"
 23757..23810
 /gene="F2J16.6"
 /note="5'-UTR sequence based on a comparison to cDNA
 L39646"
 23811..24518
 /gene="F2J16.6"
 /note="functional catalog ID=98"
 /codon_start=1
 /evidence="experimental"
 /product="ZFP3 zinc finger protein"

[illegible]

Query Match	11.4%	Score 43.8	DB 8	length 110660
Best Local Similarity	46.2%	Pred. No. 3.8		
Matches 132	Conservative 3	Mismatches 151	Indels 0	Gaps 0

QY	18	AAACCAATCTTCATTCACGAAAAATATCTTGTGGCAGCTATATTTGGTAATTTATGTCGCAAT	77
Db	85243	AAATTAATTCATGAGGATTAGATTTATTTATTCACGGCGCTCCCTACGATTTAATAATTTGT	85302

```

Oy 78 TTAATWRRRTKGGGAGATTCTCTACATGCTCTGACAGATAAAGCTCGTGCA 137
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 85303 TTTGATATATAGAAAGAAATCTTATATTAATTAATTGGACATTAACAATCTCCAA 85362
Oy 138 AATATTTGATTTATTTATTTGGAGACTTCTTATCAAAAGTAATGCTGCCAAGAGATCTAA 197
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 85363 AATGCTCTACCTTACCTACACGATTTAATTAATCTAAATTAATAAAGTCCAAATATTATA 85422
Oy 198 GGAATTACTAGTGTCCCMTCACCTGTTTGGAGTGTGCTATCTAAAGATTTTGATTC 257
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 85423 TTATTTATTTACTTCAAAATATATCATTTTGTATATGTTTCTTACATGATTAATACAT 85482
Oy 258 CTGGAATGACATATATATTTAACTTGGCTGGGGGAANAAGTTATA 303
Db 85483 CAATCCGATAGATATTTGATAGCATTTTAATTACTACAAATTTACA 85528

RESULT 11
HSU209G1
LOCUS HSU209G1 36823 bp DNA PRI 29-OCT-1997
DEFINITION Human DNA sequence from cosmid U209G1 on chromosome X.
ACCESSION Z68873
NID g1164913
VERSION Z68873.1 GI:1164913
KEYWORDS X.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Vertebrata; Mammalia; Euthera;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 36823)
AUTHORS Odell,C.
TITLE Direct Submission
COMMENT Submitted (22-JUN-1996) E-mail enquires: humquery@sanger.ac.uk
IMPORTANT: This sequence is the entire insert of clone cU209G1.
The true left end of clone U209G1 is at 1 in this sequence. The
true right end of clone V1164A6 is at 7813.
The true right end of clone U209G1 is at 36823.
This sequence was generated from part of bacterial clone contigs of
human chromosome X, constructed by the Sanger Centre chromosome X
mapping group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrx/
U209G1 is from the Lawrence Comdlibmore National Laboratory
flow-sorted X chromosome cosmid library L10XNC01.

FEATURES
Source
Location/Qualifiers
1..36823
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="U209G1"
/clone_1bp="L10XNC01"
768..853
/note="match: multiple ESTs"
1979..2268
/note="Alu repeat: matches 1..308 of consensus"
2269..2310
/note="21 copies of 2 mer 81 & conserved"
5628..7067
/note="12 copies of 120 mer 86 & conserved"
7294..7443
/note="2 copies of 75 mer 82 & conserved"
7822..7971
/note="2 copies of 75 mer 86 & conserved"
8041..8190
/note="2 copies of 75 mer 83 & conserved"
8118..8333
/note="2 copies of 108 mer 81 & conserved"
8597..8821
/note="3 copies of 75 mer 81 & conserved"
10080..10502
/note="match: 5' EST R79567 clone 145988"
complement(10976..11105)

```

```

repeat_region      /note="3 copies of 28 mer 82 & conserved"
                   30879..31173
                   /partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region      32683..32857
                   /partial
/note="Alu repeat: matches 308..133 of consensus"
repeat_region      32860..32906
                   /partial
/note="Alu repeat: matches 47..1 of consensus"
repeat_region      34532..35401
                   /note="L1 element fragment"
repeat_region      35582..35988
                   /note="L1 element fragment"
repeat_region      36039..36219
                   /note="L1 element fragment"
repeat_region      36267..36563
                   /partial
/note="Alu repeat: matches 308..1 of consensus"
BASE COUNT      10072 a 7172 c 8356 g 11223 t
ORIGIN

```

Query Match	11.4%	Score 43.8:	DB 10;	Length 36823;
Best Local Similarity	51.0%;	Pred. No. 5.1;		
Matches 102;	Conservative 0;	Mismatches 98;	Indels 0;	Gaps 0;
QY 113	TCGACAGATTAAATGTCGTGTCGCAAAATTTTGTATTTTATTTGGAGACTCTTATCAAA	172		
Db 31264	TAAACAGATTAGACACTAAATTTTAAACACACAAATTCACAGTTAAGATCTTATATAGA	31323		
OY 173	AGTATGCTGCCAAGACGACTAGAAATTAGTAGTCCCTCCTCACTTGTGGAGTG	232		
Db 31324	AATATATATAAAAAATTAAGATGACAGAGATTGAAACAGTGCAGTGAAGATATTTAACTT	31383		
QY 233	TGCTATTTCTAAAGATTTTGTATTCCTGCAATGACAATATATTTTAACTTTGGTGGGG	292		
Db 31384	TAAATTTTAAATTTAATGTCGTTTGTGTGAGTAAACAGTATATTTTAAATTAGAGATG	31443		
OY 293	AAANAGTTATAGACACAG	312		
Db 31444	GGCCAGATGATGGACATAG	31463		
RESULT 12				
LOCUS	HS164F3			
DEFINITION	Homo sapiens chromosome X clone 164F3, WORKING DRAFT SEQUENCE, in			
ACCESSION	AL035422			
NID	G5051831			
VERSION	AL035422.11			
KEYWORDS	HTG; HTGS_PHASE1.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 216591)			
JOURNAL	Wilson, S.			
COMMENT	Direct Submission Submitted (11-JUN-1999) Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humgen@esanger.ac.uk Clone requests: clonerequest@esanger.ac.uk On Jun 12, 1999 this sequence version replaced gi:5050961. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E. coli, yeast, vector, phage etc. Order of segments is not known: 800 n's separate segments. Unfinished: dj164f3 Contig_ID: 03649 acc-AL035422 Length: 216591 bp. * NOTE: This is a 'working draft' sequence.			

FEATURES
source

gene	gene
gene	/gene="F24L7.2" <4325. .>6060 /gene="F24L7.2" /note="r21L14.22; predicted by geneFeature" join(4325. .4819,4961. .4993,5123. .5175,5266. .5305, 5403. .5480,5590. .5697,5791. .5870,5967. .6060) /gene="F24L7.2" /note="hypothetical protein" /codon_start=1 /protein_id="AAC04480.1" /db_xref="pid:92914690" /db_xref="gi:2914690" /translation="WSTFDPNPNPNPNPSVSVASTPIVTASPEPTQNTVTPPSQ POPOPAPASSTAIAPLNRHHPHONIYNLPLEIRKNSVTNSPHOPPHDPSSSLTPG SSGRFPPTSPRONMSVADVGPSPGGYPRPGVYSPGHGQVSNLDMQFMRA HONOSOLSGLSHKKGVPPELRLVLYSENVTPEPTSTLSDSGKKRSRD ALVYRKRKRYITGGASYSICRSMTLNGAEGIKRIDMTCLPKPLPVKTEFSLPK DLVEALICEEDKEDEBSVKHLSBSDLKRHRIDRAKKVRLRERLRKIRYARAL LTPGEGCCRNK." complement(6589. .6697) /note="exon predicted by xgfall, quality excellent shadowexon" 6796. .6918 /note="exon predicted by xgfall, quality marginal" complement(6828. .7013) /note="exon predicted by xgfall, quality marginal shadowexon" 7708. .7972 /note="exon predicted by xgfall, quality excellent" complement(7752. .7844) /note="exon predicted by xgfall, quality excellent shadowexon" complement(7914. .7992) /note="exon predicted by xgfall, quality excellent shadowexon" complement(8108. .8225) /rpl_family="r(AAA)n" complement(8448. .8553) /note="exon predicted by xgfall, quality excellent shadowexon" complement(9131. .9157) /rpl_family="Ar_rich" complement(9372. .9408) /note="exon predicted by xgfall, quality marginal shadowexon" join(<9504. .9807,9985. .10457,10556. .>11407) /gene="F24L7.3" <9504. .>11407 /gene="F24L7.3" /note="r21L14.23" join(9504. .9807,9985. .10457,10556. .11407) /gene="F24L7.3" /codon_start=1 /product="putative phosphate transporter" /protein_id="AAC04481.1" /db_xref="pid:92914691" /db_xref="gi:2914691" /translation="MAKRGVNLNDAKQTHYHFAIYVAGSGFTDADYDFSLT VTKLGRITVHDSKSKPGTLPVNVAAVNVAVACGLDGLFPGMLGDLKGRKVG ITLMAVAVCSLGSISFGHSANGVATLCEFRWLGIGDPLSATINSEYANKRT KATLAAVAFAMSGFGLAGGIVSLVSSTPDHAKAPVTEVPGSTVPOADVYMRV LAFGALPILLYVYRMKMPETARATLAVARNTRKQASDMSKVDVLDIAEEOANSN SNSNPETGGLTFREFPARRHGHLGTTTTFLLDIAVSSSLPQKDIYTAIGMPAAE TMAHLEHPTYSKOTLIALCGTVAGWFTVAFDIDGREFIOLMGRFTMTYEMALA IPDYHMRNRNRIGFLIYSLTFMFPFANGPNAITFVPAELFPARLSTHGISASG RGAIVAGFAGFLYAQSSDSKRTDAGYPPGIGVANSLLMLACVNFGLVETLLVPESK GSLSEIISREDEBESGGSTVEMTVANSGRKVPV" 11653. .11702 /note="exon predicted by xgfall, quality marginal" complement(12069. .12123) /note="exon predicted by xgfall, quality marginal shadowexon" complement(<12375. .>13376) mRNA

Result No.	Score	Query Match	Length	DB	ID	Description
1	380	99.2	383	1	V58693	Prostate tumour sp
2	380	99.2	383	1	V61268	cDNA sequence of p
3	40	10.4	110000	1	X20248_08	Continuation (9 o
4	39.8	10.4	110000	1	V21209_03	Continuation (4 o
5	39.4	10.3	110000	1	X20248_06	Continuation (7 o
6	39.4	10.3	111309	1	X20250	Borrelia burgdorferi
7	38.6	10.1	2287	1	V22738	Babesia microti B
8	37.6	9.9	4093	1	V04419	Staphylococcus au
9	37.6	9.8	2636	1	V01871	Human brain speci
10	37.2	9.7	685	1	X19427	Nicotiana tabacum
11	36.8	9.6	110000	1	V21209_07	Continuation (8 o
12	36	9.4	2418	1	O27886	p. falciparum GAP1
13	36	9.4	3607	1	T89990	Cryptosporidium p
14	35.8	9.3	110000	1	V21209_05	Continuation (6 o
15	35.6	9.3	9047	1	X12949	Enterococcus faecoc
16	35.2	9.2	731	1	O21036	Region specific t
17	35.2	9.2	5064	1	V69110	Nelisseria meningitidis
18	35.2	9.2	400	1	V69123	Nelisseria meningitidis
19	35	9.1	8371	1	T66909	Human neuropeptid
20	35	9.1	110000	1	V21209_00	Methanococcus jann
21	35	9.1	6600	1	V86185	EST clone J635. Nae
22	35	9.1	14752	1	X20256	Borrelia burgdorferi
23	34.8	9.1	10182	1	V74427	Staphylococcus au
24	34.6	9.0	5534	1	O35988	Tomato hsp80 genom
25	34.6	9.0	2334	1	V09695	M. hyopneumonialis
26	34.6	9.0	18359	1	X20255	Borrelia burgdorferi
27	34.6	9.0	9842	1	X20259	Continuation (5 o
28	34.4	9.0	110000	1	T58840_4	Candida Cathol gen
29	34.4	9.0	3198	1	T92702	H. pylori ORF 06ep
30	34.4	9.0	3198	1	T92865	R. pilori cytoplast
31	34.4	9.0	1158	1	V24770	H. pylori GHPD 427
32	34.4	9.0	1146	1	V25046	Continuation (6 o
33	34.4	9.0	1350	1	X14148	Continuation (6 o
34	34.4	9.0	110000	1	X20248_05	Borrelia burgdorferi
35	34.4	9.0	110000	1	X20248_05	Continuation (6 o
36	34.4	9.0	116277	1	X20249	Borrelia burgdorferi
37	34.2	8.9	2749	1	T66652	Housefly Hermes e
38	34.2	8.9	3164	1	V26159	Swinepox virus Hlt
39	34.2	8.9	14752	1	V33193	Secreted protein C
40	34.2	8.9	14752	1	X20256	Borrelia burgdorferi
41	34	8.9	5852	1	O11710	Dicystostellium plas
42	34	8.9	2503	1	O53480	PMPX30 xylanase c
43	34	8.9	19124	1	T72862	Plasmodium var-7 g

	RESULT	1
ID	V58693	
AC	V58693 standard; cDNA; 383 BP.	
DT	08-DEC-1998 (first entry)	
DE	Prostate tumour specific gene clone P509S.	
KW	Prostate tumour specific gene; human; prostate cancer; detection;	
RN	therapy; ss.	
OS	Homo sapiens.	
PN	M09837418-A2.	
PF	27-AUG-1998.	
PD	25-FEB-1998; U03690.	
PR	09-FEB-1998; US-904809.	
PR	25-FEB-1997; US-806596.	
PR	01-AUG-1997; US-904809.	
PA	(CORI-) CORIXA CORP.	
PI	DILLON DC, XU J,	
WP:	98-480805/41.	
PT	Novel human prostate specific tumour protein and fragments - useful	
PT	for detecting and treating prostate cancers	
PS	Claim 1; Page 134; 141pp; English.	
CC	This sequence represents a human prostate tumour specific gene, and can	
CC	be used in the method of the invention. The method is for detecting	
CC	prostate cancer comprises contacting a biological sample with an agent	
CC	able to bind an immunogenic portion of a prostate protein (such as	
CC	encoded by this sequence). An antibody which binds to an immunogenic	
CC	portion of the prostate protein, and the method can be used to detect,	
CC	monitor progression of, or treat prostate cancers. The antibody may	
CC	also be conjugated to a therapeutic agent for use in therapy of prostate	
CC	cancers.	
SQ	Sequence 383 BP; 123 A; 51 C; 63 G; 140 T;	

Db 41472 CAAATTCGCTAAGATGCTTGGTTAGAGCTCAGATGATGCTTCAAGAGTTGC 41413
 Oy 181 TGGCAAGAGAGTCAAGATGATGATGCT 211
 Db 41412 ATCCAAAGTTCGCTCAGATTAATCAAGGT 41382

RESULT 5

X20248_06
 Continuation (7 of 10) of X20248 from base 600001 (Borrelia burgdorferi polynucleotide s
 WP Sequence split into 10 fragments LOCUS X20248 Accession X20248

Fragment Name	Begin	End
WP X20248_00	1	110000
WP X20248_01	100001	210000
WP X20248_02	200001	310000
WP X20248_03	300001	410000
WP X20248_04	400001	510000
WP X20248_05	500001	610000
WP X20248_06	600001	710000
WP X20248_07	700001	810000
WP X20248_08	800001	910000
WP X20248_09	900001	910715

Query Match 10.3%; Score 39.4; DB 1; Length 110000;

Best Local Similarity 43.5%; Pred. No. 1.5; 196; Indels 0; Gaps 0;

Matches 154; Conservative 4; Mismatches 196; Indels 0; Gaps 0;

Oy 29 TCATTCAGAAAAATTAATCTTAGGAGCTGATTTGTAATATGTCATTTAATWRTT 88
 Db 76486 TGATTCCTTAATCTTATGTTAAAGACTTAAGCTTAATGATTAATTAAGTTT 76545
 Oy 89 KTGGGGCATTTCCTTACATGCTCTTGACAGATTAAGTCTGTGCCAAAATTTGTAT 148
 Db 76546 TCCATCGATTCCTCGTCGAGCTTTCAGATATTTATTTGCCCCAAAGTATTTGGAT 76605
 Oy 149 TTTATTTGGAGACTTCTTATCAAAAGTAATGCTGCCAAGAGCTTAAGCAATAGTAG 208
 Db 76606 TTTATTTTATGTTATGTTAGTAATATTTTGTGATTAATCTTGAGAGCTCTTTT 76665
 Oy 209 TGTTCCTCCTGCTGTTGGAGTGTCTATCTTAAGATTTGATTTCCGTGAATGACA 268
 Db 76666 GAATTAAGATAGATATGATATCTTCTTTTCTAAAAATCTAATTTAAAAATTAATT 76725
 Oy 269 ATTAATATTTAACTTGTGGTGGGGAANAAGTATAGACACAGCTCTCACTTGATAC 328
 Db 76726 TACATTCGAAATATGTTACTTCTTAACCTTAAGAAATGATGATTTAGCTTTTATTTTC 76785
 Oy 329 TTGTAATTAATCTTTATGCACTGTTTGACCTAATAGCTATATGTTTAA 382
 Db 76786 ATTAAGATCCAACTTTTCTATTTGTTTGGCAAACTTCTAATTTTATAGTAA 76839

RESULT 6

X20250
 ID X20250 standard: DNA; 111309 BP.

AC X20250;
 DT 04-MAY-1999 (first entry)
 DE Borrelia burgdorferi polynucleotide sequence #3.
 KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 KM epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
 OS Infection; diagnosis; characterisation; detection; ds.
 PN Borrelia burgdorferi.
 PN W09858943-A1.
 PD 30-DEC-1998.
 PF 18-JUN-1998; U12764.
 PR 03-SEP-1997; US-057483.
 PR 20-JUN-1997; US-050359.
 PR 22-JUL-1997; US-053344.
 PR 22-JUL-1997; US-053377.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI (MED-) MEDIMUNE INC.
 PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,

PI White OR:
 DR WPI: 99-081217/07.

PT New isolated Borrelia burgdorferi nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of infections, particularly Lyme disease
 PS Claim 1: Page 738-800; 1128pp; English.
 CC X20248 to X20402 represent polynucleotide sequences isolated from
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
 CC the detection, diagnosis, characterisation, prevention and therapy of
 CC Bb infections, e.g. Lyme disease. They can also be used for the
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
 CC Lyme disease.
 CC Sequence 111309 BP; 35956 A; 13151 C; 19075 G; 43117 T;

Query Match 10.3%; Score 39.4; DB 1; Length 111309;

Best Local Similarity 43.5%; Pred. No. 1.5; 196; Indels 0; Gaps 0;

Matches 154; Conservative 4; Mismatches 196; Indels 0; Gaps 0;

Oy 29 TCATTCAGAAAAATTAATCTTAGGAGCTGATTTGTAATATGTCATTTAATWRTT 88
 Db 87100 TGATTCCTTAATCTTATGTTAAAGACTTAAGCTTAATGATTAATTAAGTTT 87159
 Oy 89 KTGGGGCATTTCCTTACATGCTCTTGACAGATTAAGTCTGTGCCAAAATTTGTAT 148
 Db 87160 TCCATCGATTCCTCGTCGAGCTTTCAGATATTTATTTGCCCCAAAGTATTTGGAT 87219
 Oy 149 TTTATTTGGAGACTTCTTATCAAAAGTAATGCTGCCAAGAGCTTAAGCAATAGTAG 208
 Db 87220 TTTATTTTATGTTATGTTAGTAATATTTTGTGATTAATCTTGAAGACCTCTTTT 87279
 Oy 209 TGTTCCTCCTGCTGTTGGAGTGTCTATCTTAAGATTTGATTTCCGTGAATGACA 268
 Db 87280 GAATTAAGATAGATATGATATCTTCTTTCTAAAAATCTAATTTAAAAATTAATT 87339
 Oy 269 ATTAATATTTAACTTGTGGTGGGGAANAAGTATAGACACAGCTCTCACTTGATAC 328
 Db 87340 TACATTCGAAATATGTTACTTCTTAACCTTAAGAAATGATGATTTAGCTTTTATTTTC 87399
 Oy 329 TTGTAATTAATCTTTTATGCACTGTTTGACATTAAGCTATATGTTTAA 382
 Db 87400 ATTAAGATCCAACTTTTCTATTTGTTTGGCAAACTTCTAATTTTATAGTAA 87453

RESULT 7

V22738/C
 ID V22738 standard: DNA; 2287 BP.

AC V22738;
 DT 28-SEP-1998 (first entry)
 DE Babesia microti BMNI-8 antigen sequence.
 KW antigen; detection; diagnosis; vaccine; tick-borne disease;
 KM differentiation; Lyme disease; ehrlichiosis; ss.
 OS Babesia microti.
 FH key Location/Qualifiers
 FT CDS 31..1806 /tag= a
 FT /product= antigen
 PN BP-834567-A2.
 PD 08-APR-1998.
 PF 01-OCT-1997; 117067.
 PR 24-APR-1997; US-845258.
 PR 01-OCT-1996; US-723142.
 PA (CORI-) CORIAX CORP.
 PI Houghton R, Lodes MJ, Reed SG, Sleath PR;
 DR WPI: 98-193465/18.
 DR P-PDSB; W56287.
 PT Polypeptides comprising Babesia microti antigens and their
 PT immunogenic fragments or epitopes - and related nucleic acid,
 PT vectors, transfected cells and antibodies, useful for diagnosis of
 PT infection and in protective vaccines

```

RESULT      9
VO1871/c
ID      VO1871 standard; DNA; 2636 BP.
AC      VO1871;
DT      20-APR-1998 (first entry)
DE      Human brain specific nucleosome assembly protein gene.
KW      Human; fetal brain cDNA library; GDP dissociation stimulating protein
KW      brain specific nucleosome assembly protein; diagnosis; therapy;
KW      skeletal muscle specific ubiquitin conjugating enzyme; TMP-2; NPK;
KW      nel-related protein type 1; nel-related type 2; hereditary disease;
KW      cancer; ss.
OS      Homo sapiens.
FH      key
FH      CDS
FT      Location/Qualifiers
FT      266..1786
FT      /tag= a
FT      /product= "brain specific nucleosome assembly protein"
EP-796913-A2.
PN      24-SEP-1997.
PD      19-MAR-1997; 104842.
PF      05-MAR-1997; JP-069163.
PR      19-MAR-1996; JP-063410.

```

PA (SAKURA) OTSUKA PHARM CO. LTD.
PI Fujiwara T, Horie M, Matsumae T;
DR WPI: 97-459830/43.
DR P-PSDB; W374455.
PT Novel human genes, e.g. brain-specific nucleosome assembly protein -
PT useful for diagnosis or therapy of hereditary disease and cancer
PS Claim 6; Page 60-64; 123pp; English.
CC The present sequence encodes a brain specific nucleosome assembly
CC protein isolated from a human foetal brain cDNA library. The nucleotide
CC or amino acid sequences are useful for in-vitro diagnosis of hereditary
CC diseases and cancer and for preparation of pharmaceuticals.
SO Sequence 2636 BP; 814 A; 521 C; 599 G; 702 T;

Query Match 9.88; Score 37.6; DB 1; Length 2636;
Best Local Similarity 46.68;
Matches 109; Conservative 3; Mismatches 122; Indels 0; Gaps 0;

[illegible]

RESULT 10
ID X19427
AC X19427 standard; DNA; 685 BP.
DT 04-JUN-1999 (first entry)
DE Nicotiana tabacum matrix attachment region ps211-1 DNA sequence.
KW Nicotiana tabacum; matrix attachment region; MAR; tabacco;
expression; transgene; structural gene; ss.
OS Nicotiana tabacum.
PN W09907866-A1.
PD 18-FEB-1999.
PF 05-AUG-1998; U16344.
PR 06-AUG-1997; US-066118.
PT (UYNC-) UNIV NORTH CAROLINA STATE.
PA Michalowski SM, Spiker SL.
PI WPI; 99-167436/14.
PT New Matrix Attachment Region DNA - useful for increasing expression
of transgene product, especially structural genes in tobacco
PS Claim 1; Fig 3: 66pp; English.
CC The present sequence represents DNA encoding a Matrix Attachment Region
CC (MAR) (which bind to the nuclear matrix) isolated from Nicotiana tabacum
CC A DNA construct containing MAR sequences can be used to increase
CC expression of a transgene product. A new method identifies MARs by
CC sequence alone, unlike prior art methods using associated genes. MARs
CC increase expression of a transgene product.
SQ Sequence 685 BP; 243 A; 107 C; 87 G; 248 T;

Query Match	9.7%	Score	37.2	DB	1	Length	665
Best Local	Similarly	49.4%					
Matches	87	Conservative	3	Mismatches	86	Indels	0
						Gaps	0

QY 31 ATTCGAAAATTTACTTAGGAGCTGATATTTGGTAATTATCGCAATTTAAATWTRTKT 90
 Db 366 ATTAATAAATAATAGATCGGGCGCTGCATTTTATATTAATATTTATTTATTTT 425
 QY 91 GGGGCATTCTTACATTGTCTTGACAAGATTAAATGCTGTGCCAAAATTTGTATTT 150

Db 426 AGCGAAGTCCCTCCCTATTTTATGTAATACCCCTTTAATGACACACTTATTATTACT 485

Qy 151 TATTTCGAGACTTCTTATCAAAATATATGCTGCCAAGAAGCTTAAGAAATTACT 206

Db 486 AAGTTTGCTAATTATATGAAGTCAATCTCAGTACATATAAATAATACATATTAT 541

RESULT	11
V21209_07	
Continuation (8 of 17) of V21209 from base 700001 (Methanococcus jannaschii circular	
WP	WP Sequence split into 17 fragments
WP	Fragment Name
WP	V21209_00
WP	V21209_01
WP	V21209_02
WP	V21209_03
WP	V21209_04
WP	V21209_05
WP	V21209_06
WP	V21209_07
WP	V21209_08
WP	V21209_09
WP	V21209_10
WP	V21209_11
WP	V21209_12
WP	V21209_13
WP	V21209_14
WP	V21209_15
WP	V21209_16
	Begin
	1
	End
	110000
	210000
	310000
	410000
	510000
	610000
	710000
	810000
	910000
	1010000
	1110000
	1210000
	1310000
	1410000
	1510000
	1610000
	1664976
	Accession V21209

Query Match	9.6%	Score 36.8;	DB 1;	Length 110000;
Best Local Similarity	45.2%			
Matches 122; Conservative	3;	Pred. No. 5.7;	145;	Indels 0;
		Mismatches		Gaps

QY 18 AAGCAATTCCTCATGCANAAAAATATCTGAGGACGACGATATGGTAATATGCGCAAT 77
Db 64033 ATAGTATATTTACACTTTTATATTTATTTTGGATTAATATTAATTAATTTTTTAAT 64092
QY 78 TTAATWTRTTRTKTGSGGCAATTCCTTACATGTCTTGACAGATTAAATGTGTGCGCA 137
Db 64093 ATTTTCGACATTTAATATTTTAAATGTAATTTTTTTGGCAAAAAAAGAACAAATTTTCCTG 64152
QY 138 AATTTTGTATTTTATTGGAACCTCTTATCCAAAAGTAAGCGCCAAAGGAAGCTAA 197
Db 64153 AATTTTAAATGAATTTTAAAGAAAAAAGAAAAAATATCTTAATTAATAATTAATAAT 64212
QY 198 GGAATTAGTGTCTCCCTCACTGTTTGAGAGTGCTATCTCAAAAAGATTTGATTTTC 257
Db 64213 TAGAAATTAAGTAAATTCATATTTTATTTGATTTAAATTTTACCTATCCCTTATTTAT 64272
QY 258 CTGGATGACAAATTAATTTTAACCTTGCT 287
Db 64273 TCCTTTTGACAAATTTTTCACATTTTGT 64302

RESULT	12
027886/c	
ID	Q27886 standard; DNA; 2418 BP.
AC	Q27886;
DT	04-FEB-1993 (first entry)
DE	P. falciparum GBP10h.
KW	Polymerase chain reaction; glycoporin binding protein; Inverse PCR.
OS	Plasmodium falciparum; malaria; ss.
FS	synthetic.
FH	Key
FT	Location/Qualifiers
FT	767..955
FT	/*tag= a
FT	/number= 1
FT	1111..2202
FT	exon
FT	/*tag= b
FT	/number= 2
FT	1249..2202
FT	repeat-region

FT EP-498934-A. /tag- c
 PN 26-AUG-1992.
 PD 27-JAN-1992: 101271.
 PF 21-FEB-1991: DE-105348.
 PR (BEHN) BEHRINGER AG.
 PI Hundt E, Knapp B, Kupper H, Nolte D, Kuepper H;
 DR WPI: 92-286009/35.
 PT DNA coding protein GBP 130 h and proteins produced - used for
 PS vaccines to control malaria
 CC Claim 1: Page 12-14; 17pp; German.
 CC Glycophorin binding protein 130h is a P.falciparum blood stage
 CC antigen 69% homologous with the known GBP130. The GBP130h gene and
 CC its homologues can be used in vaccines to protect against malaria
 CC caused by P.falciparum. The coding sequence was isolated using PCR
 CC techniques on a genomic P.falciparum FCBR strain library. The
 CC GBP130h gene was found to be highly conserved between different
 CC strains but is distinct from the GBP130 gene.
 CC See Q27878-Q27885 and Q33432.
 SQ Sequence 2418 BP; 1047 A; 303 C; 292 G; 776 T;

Query Match 9.4%; Score 36; DB 1; Length 2418;
 Best Local Similarity 42.6%; Pred. No. 5.6; 210; Indels 0; Gaps 0;
 Matches 159; Conservative 4; Mismatches 210; Indels 0; Gaps 0;

QY 1 AAACAACAACAAAAAACAATTCCTTCATTCAGAAAAATTAATCTTAGGAGCTGATAT 60
 DB 638 ATAAAAAATAAAAAAACAATACATAATATTTTATTTATTTATTTATTTATTTT 579
 QY 61 TGGTAATATGTCATATTAATWRTTKTGGGCAATTCCTTACATGCTTGACAGA 120
 DB 578 TTTTTCATTAATAATTAATACATGTTTTCATTAATAAAGATAATATATAT 519
 QY 121 TTAATATGCTGCGCAAAATTTGATTTTATTTGAGACTTCTTCAAACTATGC 180
 DB 518 ATATTTTCAAGATTAATTAATATTCATTCCTGAAAAGATATATATATTTATAT 459
 QY 181 TGCACAGAGAGCTAGAGATTAAGTATGTCCTCCATCTGTTGAGATGCTATTC 240
 DB 458 TAGTAGAGTACAGAGAGCAGAAATATATATGATATTAACCTAGACATTTGATATA 399
 QY 241 TAAAGATTTGATTTCCGGAATGACATTAATTTTAACTTTGGTGGGGAANAAGT 300
 DB 398 AATATTAATTTATTTATGATATATATTTTATTTGATCTTAAATTAATAAGAA 339
 QY 301 ATAGGACAGAGCTTCATCTGATCTGTAATTAATCTTTATGACCTGTTTG 360
 DB 338 TTAGAAATATATTTTATATATATATTTTCCACATGTTCTTTGATCTATCTTTTC 279
 QY 361 ACCATTAAGCTAT 373
 DB 278 TATTATAAATAT 266

RESULT 13
 T89990/c
 ID T89990 standard: DNA; 3607 BP.
 AC T89990:
 DT 11-MAY-1998 (first entry)
 KW Cryptosporidium parvum oocyst heat shock protein HSP70 gene.
 KM Cryptosporidium detection; assay; protozoan; heat shock protein;
 OS HSP70; gene amplification; PCR; ds.
 OS Cryptosporidium parvum.
 FH Key location/Qualifiers
 FT CDS 1272..3296
 FT /tag- a
 FT /transl_except- (pos:1536..1538; aa:His)
 FT /transl_except- (pos:1578..1580; aa:Ile)
 FT /transl_except- (pos:1659..1661; aa:Lys)
 FT /transl_except- (pos:2034..2036; aa:Lys)
 FT /transl_except- (pos:2265..2267; aa:Leu)
 FT /transl_except- (pos:2931..2933; aa:Lys)

FT /transl_except- (pos:3093..3095; aa:Lys)
 FT /transl_except- (pos:3105..3107; aa:His)
 FT /transl_except- (pos:3267..3269; aa:Lys)

MO9742349-A1.

PN 13-NOV-1997.
 PD 08-MAY-1997; 007972.
 PF 09-MAY-1996; US-647351.
 PA (METR-) METROPOLITAN WATER DISTRICT SOUTHERN CALIFORNIA.
 PI De Leon R, Rochelle PA;
 DR WPI: 97-556999/51.
 P-PSDB; W31345.

PT Amplifying and detecting viable and infective protozoan oocysts -
 PT by monitoring presence or absence of heat shock protein genes in
 PT their transcripts
 PS Claim 12: Page 23-26; 44pp; English.

CC This genomic DNA sequence comprises the heat shock protein HSP70
 CC gene of Cryptosporidium parvum oocysts. The invention provides
 CC processes and kits for detecting encysted forms of protozoa,
 CC particularly Cryptosporidium (especially Cryptosporidium muris,
 CC Cryptosporidium baileyi, Cryptosporidium wrairi and C. parvum)
 CC and Giardia, that are viable and infectious by the enzymatic
 CC amplification (e.g. by PCR) of a target gene sequence such as HSP70.
 CC A preferred primer pair (see T89991-92) can amplify DNA or RNA
 CC from many different Cryptosporidium species, and can be used with a
 CC product (see T89997) that confirms the identity of the amplification
 CC to C. parvum and can also be used with a probe (see T89998) to
 CC confirm sequence identity. In addition, both Cryptosporidium and
 CC Giardia can be detected simultaneously using a multiplex
 CC amplification reaction. The methods exploit the speed, sensitivity
 CC and specificity associated with an amplification procedure, and
 CC enable pathogenic forms of protozoa present in low copy numbers to
 CC be identified and distinguished from morphologically similar, but
 CC non-pathogenic protozoa.
 SQ Sequence 3607 BP; 1265 A; 514 C; 716 G; 1112 T;

Query Match 9.4%; Score 36; DB 1; Length 3607;
 Best Local Similarity 48.3%; Pred. No. 5.8; 89; Indels 0; Gaps 0;
 Matches 87; Conservative 4; Mismatches 89; Indels 0; Gaps 0;

QY 1 AAACAACAACAAAAAACAATTCCTTCATTCAGAAAAATTAATCTTAGGAGCTGATAT 60
 DB 221 AATAAAAAATAAAAAATACACATTTATTTAATAAATCTTAATAAATAAGTGAATA 162
 QY 61 TGGTAATATGTCATATTAATWRTTKTGGGCAATTCCTTACATGCTTGACAGA 120
 DB 161 TGGATCTCCATCAATCAAAATTAATTTTACAGACTATCAACAATATGATTAATA 102
 QY 121 TTAATATGCTGCGCAAAATTTGATTTTATTTGAGACTCTTAATCAAAAGTATGC 180
 DB 101 TGAACATAATGCTTGAATAATTAATTTTCATAGGAAAAAATTTTATTAATATTTAC 42

RESULT 14
 V21209_05
 Continuation (6 of 17) of V21209 from base 500001 (Methanococcus jannaschii circular
 WP Sequence split into 17 fragments LOCUS V21209 Accession V21209
 WP Fragment Name Begin End
 WP V21209_00 1 110000
 WP V21209_01 100001 210000
 WP V21209_02 200001 310000
 WP V21209_03 300001 410000
 WP V21209_04 400001 510000
 WP V21209_05 500001 610000
 WP V21209_06 600001 710000
 WP V21209_07 700001 810000
 WP V21209_08 800001 910000
 WP V21209_09 900001 1010000
 WP V21209_10 1000001 1110000
 WP V21209_11 1100001 1210000
 WP V21209_12 1200001 1310000
 WP V21209_13 1300001 1410000

OY 175 TAATCTGCCAAGAGTCTAAGAAATAGTGTCCCTCACTGTTGGAGCTGTG 234
DB 1919 GAAACCAACAAACTGACTATAGGAATATAGTGTGAGTTAGAGCTCAGGGCTTTA 1860
OY 235 CATATCTAAATATTTGATTTCCCTGGAATGACAAATATTTTAACTTTGGTG 288
DB 1859 CTTTTCCTCAAGGCTGTATGAATCTGCTTCACTTCTTGAGATTTTAAATCTTG 1806

RESULT 2

US-08-647-351B-1/c
; Sequence 1, Application US/08647351B
; Patent No. 5770368
; GENERAL INFORMATION:
; APPLICANT: De Leon, Ricardo
; APPLICANT: Rochelle, Paul
; TITLE OF INVENTION: Cryptosporidium Detection Method
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 225 S. Lake Avenue, 9th Floor
; CITY: Pasadena
; STATE: California
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows version 6.1
; CURRENT APPLICATION DATA: US/08/647,351B
; FILING DATE: May 9, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Farah, David A.
; REGISTRATION NUMBER: 38,134
; REFERENCE/DOCKET NUMBER: 11364
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 796-4000
; TELEFAX: (626) 795-6321
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3607 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double stranded
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-647-351B-1

Query Match 9.4%; Score 36; DB 3; Length 3607;
Best Local Similarity 48.3%; Pred. No. 0.72;
Matches 87; Conservative 4; Mismatches 89; Indels 0; Gaps 0;

OY 1 AAAACAAACAAACAAACAAACAAATTCCTCAATCAGAAAAATTAATCTTAGGAGCTGATAT 60
DB 221 AATAAATAAATAAATAAATAAATAAATTTAATAAATAAATAAATAAATAAATAAATAA 162
OY 61 TGGTAATATGCTCAATTAATATTTTGGGCAATTCCTTACATCTCTGACAGAA 120
DB 161 TGGATCTCTCTCAATCAATTAATATTTTACAGACTTAATCAACAAATATGATTTATATA 102
OY 121 TTAATATGCTGTGCCAAATTTTGTATTTTGGAGACTTCTTATCAAAAGTAATGC 180
DB 101 TGAACCTATGCTGTAAATTAATTTATATGAGAAAAAATTTTATATATATTTAC 42

RESULT 3
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:

; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526r1s
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; US-07-867-106-2

Query Match 9.1%; Score 34.8; DB 1; Length 5852;
Best Local Similarity 48.9%; Pred. No. 1.6;
Matches 87; Conservative 2; Mismatches 89; Indels 0; Gaps 0;

OY 1 AAAACAAACAAACAAACAAACAAATTCCTCAATCAGAAAAATTAATCTTAGGAGCTGATAT 60
DB 2022 AAAAAAATAAATAAATAAATAAATAAATTAATTAATTAATTAATTAATTAATTAATTA 2081
OY 61 TGGTAATATGCTCAATTAATATTTTGGGCAATTCCTTACATCTCTGACAGAA 120
DB 2082 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2141
OY 121 TTAATATGCTGTGCCAAATTTTGTATTTTGGAGACTTCTTATCAAAAGTAAT 178
DB 2142 TTGATATATATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2199

RESULT 4
US-08-062-632-4/c
; Sequence 4, Application US/08062632
; Patent No. 5712090
; GENERAL INFORMATION:
; APPLICANT: Artushin, Sergey
; APPLICANT: Stipkovits, Iaslo

APPLICANT: Minlon, F. Chis
TITLE OF INVENTION: PCR-Based Assay for Mycoplasma
TITLE OF INVENTION: Hypopneumoniae
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dickstein, Shapiro and Morin
STREET: 2101 L. St. NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 22037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/062,632
FILING DATE: 18-MAY-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brady Jr., James W.
REGISTRATION NUMBER: 32,115
REFERENCE/DOCKET NUMBER: 18900.018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)785-9700
TELEFAX: (202)887-0689
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2334 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-062-632-4

Query Match 9.0%; Score 34.6; DB 2; Length 2334;
Best Local Similarity 48.6%; Pred. No. 1.4;
Matches 85; Conservative 3; Mismatches 87; Indels 0; Gaps 0;
QY 1 AAAAAACAAACAAAAAATTCCTTCAGAAAAATTAATCTTGGGAGCTGATAT 60
DB 496 AAAAAAGGAAAGAAATATTAATAAATGATATTAATTAATTAATTAATTA 437
QY 61 TGTAAATTAATGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 120
DB 436 CAGATATAAGTCACTCTTACTTAATTAATTAATTAATTAATTAATTAATTA 377
QY 121 TTAATATGCTGTCGCAAAATTTGTAATTTGAGAGCTGCTTATCAAAAGT 175
DB 376 TTACTATGTTGCTCAATTAATTCAGACACTTTTATTAATTAATTAATTAAT 322

RESULT 5
US-08-344-695-1/c
Sequence 1, Application US/08344695
Patent No. 5614398
GENERAL INFORMATION:
APPLICANT: O'BROCHTA, DAVID
ADDRESSEE: WARREN, WILLIAM
APPLICANT: ATKINSON, PETER
TITLE OF INVENTION: A GENE TRANSFER SYSTEM FOR INSECTS
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/344,695
FILING DATE: 18-NOV-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Keiber, Steven B.
REGISTRATION NUMBER: 30,073
REFERENCE/DOCKET NUMBER: 2747-058-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TEXT: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2749 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 450..2285
US-08-344-695-1

Query Match 8.9%; Score 34.2; DB 1; Length 2749;
Best Local Similarity 49.2%; Pred. No. 1.9;
Matches 87; Conservative 1; Mismatches 89; Indels 0; Gaps 0;
QY 14 AAAAAACAAATCTTATTCAGAAAAATTAATCTTGGGAGCTGATATTAATGCT 73
DB 1147 AAAAAATGCTTAACCTTTAATAAATTTCTGCTGATCTTAATTAATTAATTA 1088
QY 74 CAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 133
DB 1087 GACTTAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1028
QY 134 GCCAAATTTGTAATTTGAGAGCTTCTATCAAAATTAATGCTGCCAAAGCA 190
DB 1027 CCAAAATTCGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 971

RESULT 6
US-07-867-106-2/c
Sequence 2, Application US/07867106
Patent No. 5389526
GENERAL INFORMATION:
APPLICANT: Slade, Martin B
ADDRESSEE: Chang, Andy C M
APPLICANT: Williams, Keith L
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526rls
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625
PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU PJ 7187
APPLICATION NUMBER: PCT/AU90/00530
FILING DATE: 02-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Feeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
US-07-867-106-2

Query Match 8.9% Score 34; DB 1; Length 5852;
Best Local Similarity 48.8%; Pred. No. 2.4;
Matches 82; Conservative 3; Mismatches 83; Indels 0; Gaps 0;

OY 1 AAAACAAACAAACAAACAAATCTTCATCAGAAAAATATCTTAGGAGCTGATAT 60
DB 5626 AAAAAAAAAAAAAAAAAAAAAAAAAAATGATATAATTTGGAATTAATAAATA 5567
OY 61 TGGTAATTATGCGCATTTATATRTTTRTTGGGCGCAATTCCTTACATGCTTGACAAGA 120
DB 5566 AATGTAAAGGTTTTTTTAAATTTATGATCGATTTTATTTAAACATTCAGACAG 5507
OY 121 TTAATAATGCTGTGCCAAATTTGTATTTTATTTGGAGACTCTTAT 168
DB 5506 TTAATAAATCTTAAACAAATTAACATATTTGATTTTTTTTTTTTTTTT 5459

RESULT 7

US-08-446-855A-1
Sequence 1, Application US/08446855A
Patent No. 5849573
GENERAL INFORMATION:
APPLICANT: Stewart, Thomas S
APPLICANT: Flores, Maria V
APPLICANT: O'Sullivan, William J
TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 1100 No. 5849573th Glebe Road, 8th floor
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,855A
FILING DATE: 06-Jul-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mltchard, Leonard C

REGISTRATION NUMBER: 29,009
REFERENCE/DOCKET NUMBER: 47-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic
US-08-446-855A-1

Query Match 8.8% Score 33.6; DB 4; Length 8920;
Best Local Similarity 59.3%; Pred. No. 3.3;
Matches 51; Conservative 3; Mismatches 32; Indels 0; Gaps 0;

OY 1 AAAACAAACAAACAAACAAATCTTCATCAGAAAAATATCTTAGGAGCTGATAT 60
DB 633 AAAAAAAAAAAAAAAAAAAAAAAAAAATTTACATATGAAAAATGAACTGTATATGTAAT 692
OY 61 TGGTAATTATGCGCATTTATATRTTTRTTGGGCGCAATTCCTTACATGCTTGACAAGA 86
DB 693 TTATAAATATTTTAAACATTAATAATA 718

RESULT 8

US-08-414-685-1
Sequence 1, Application US/08414685
Patent No. 5667989
GENERAL INFORMATION:
APPLICANT: McCullough, John
APPLICANT: Baymiller, Judy
TITLE OF INVENTION: Fungal Cell Wall Protein CLY4
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burton Rodney
STREET: P.O. Box 4000
CITY: Princeton
STATE: New Jersey
COUNTRY: USA
ZIP: 08543-4000
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/414,685
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Gaul, Timothy J.
REGISTRATION NUMBER: 33,111
REFERENCE/DOCKET NUMBER: DC33
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 252-5901
TELEFAX: (609) 252-4526
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2074 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 354..1193
US-08-414-685-1

RESULT 11
US-08-832-883-65/c
; Sequence 65, Application US/08832883
; Patent No. 5807681
; GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; APPLICANT: Baldi, Alphonso
; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS
; TITLE OF INVENTION: OF CANCER
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEIDEL, GONDA, LAVORONA & MONACO, P.C.
; STREET: Suite 1800 Two Penn Center Plaza
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/832,883
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-13 US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 901 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-883-65

Query Match 8.1%; Score 31.2; DB 3; Length 901;
Best Local Similarity 47.1%; Pred. No. 8.2; Mismatches 122; Indels 2; Gaps 1;
Matches 114; Conservative 4;
QY 20 AACATTCCTTCATTCAGAAAATTAATCTTAGGACGATATGTAATTAATGTCATTT 79
DB 774 AATAACTTAACCTAAAGAAAATATCTAATAGCAACCTAAGTGTGTTGTTGCCA 715
QY 80 AATWRTTRTKGGGGCATTCCTTACAT--TGCTTGACAAGATTAAAGTCTGTCGA 137
DB 714 GTTATTAATGTAAGCAATTTTATACATTTATGTCATGTCATCCCTACAAACCCCATAT 655
QY 138 AATTTTGTATTTATTTAGAGACTTCTTATCAAAAGTAATGCTGCAAGGAAGCTAA 197
DB 654 GCTAGAACACTAGTTATATTCCTCATTTTATATATAGAGAAAATGAGTACAGAAAATTTA 595
QY 198 GGAATTAGTAGTGTCCCMTCACCTGTTGGAGTGTGCTATTCATAAAGATTGATTTC 257
DB 594 ATGATTTGACGGGTTTACACACACTGTAAAGTACGGAAATTTGAACCCATGTAGCCTGC 535
QY 258 CT 259
DB 534 CT 533

RESULT 12
US-08-832-877-65/c
; Sequence 65, Application US/08832877
; Patent No. 5840506

GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS OF
; TITLE OF INVENTION: CANCER
; NUMBER OF SEQUENCES: 116
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEIDEL, GONDA, LAVORONA & MONACO, P.C.
; STREET: Suite 1800 Two Penn Center Plaza
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/832,877
; FILING DATE:
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-13 US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 901 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-877-65

Query Match 8.1%; Score 31.2; DB 4; Length 901;
Best Local Similarity 47.1%; Pred. No. 8.2; Mismatches 122; Indels 2; Gaps 1;
Matches 114; Conservative 4;
QY 20 AACATTCCTTCATTCAGAAAATTAATCTTAGGACGATATGTAATTAATGTCATTT 79
DB 774 AATAACTTAACCTAAAGAAAATATCTAATAGCAACCTAAGTGTGTTGTTGCCA 715
QY 80 AATWRTTRTKGGGGCATTCCTTACAT--TGCTTGACAAGATTAAAGTCTGTCGA 137
DB 714 GTTATTAATGTAAGCAATTTTATACATTTATGTCATGTCATCCCTACAAACCCCATAT 655
QY 138 AATTTTGTATTTATTTAGAGACTTCTTATCAAAAGTAATGCTGCAAGGAAGCTAA 197
DB 654 GCTAGAACACTAGTTATATTCCTCATTTTATATATAGAGAAAATGAGTACAGAAAATTTA 595
QY 198 GGAATTAGTAGTGTCCCMTCACCTGTTGGAGTGTGCTATTCATAAAGATTGATTTC 257
DB 594 ATGATTTGACGGGTTTACACACACTGTAAAGTACGGAAATTTGAACCCATGTAGCCTGC 535
QY 258 CT 259
DB 534 CT 533

RESULT 13
US-08-480-604A-9
; Sequence 9, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.

```

Patent No. 5538525
GENERAL INFORMATION:
APPLICANT: BROEKERT, WILLEM F.
APPLICANT: CAMOIE, BRUNO P.A.
APPLICANT: OSBORN, ROBERT W.
APPLICANT: REES, SARAH B.
APPLICANT: TERRAS, FRANKY R.G.
APPLICANT: VANDERLEIDEN, JOZEF
TITLE OF INVENTION: BIOCIDAL PROTEINS
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DABY & CUSHMAN
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/377,687
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/002,480
FILING DATE: 04-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: KORULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 99042/SEE.36525/US/A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 414 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 16..255
US-08-377-687-48

Query Match 8.0%; Score 30.6; DB 1; Length 414;
Best Local Similarity 46.9%; Pred. No. 9.8;
Matches 84; Conservative 3; Mismatches 92; Indels 0; Gaps 0

OY 20 AACATTTCTTCATTCAGAAAATTAATCTTAGGAGCTGATTTGGTAATATGCTCAATTT 79
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 391 AATAAAATTTATTTGTTAATACCGAACCAACATTAATCTGGTACATGCTTTAAATC 332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 80 AATWATTTTKGGGCGCATTCCTTCATTTGCTTGTCGCAAGATTAAATGTCGTGGCAAA 139
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 331 ACTCATGGATGTGTCACATGACTTATTTTGTGTAATTAATTCATAAAACATTAATTCACCAAG 272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 140 ATTTTGATTTTATTTTGAGACTTCTTATTCAAAGTAATGCTGCCAAGGAAGTCTAAG 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 271 AGTTTGCATTAATTAACAGAGGAAGTAGCATATCATCTGTGAGCTGGGAAGACATAG 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
US-08-206-176-3
: Sequence 3, Application US/08206176
: Patent No. 5639940
: GENERAL INFORMATION:
: APPLICANT: Garner, Ian
: APPLICANT: Dalrymple, Michael A

```

```

APPLICANT: Prunkard, Donna E
APPLICANT: Foster, Donald C
TITLE OF INVENTION: Production of fibrinogen in Transgenic
TITLE OF INVENTION: Animals
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: ZymoGenetics, Inc.
STREET: 4225 Roosevelt Way, N.E.
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/206,176
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E
REGISTRATION NUMBER: 31-648
REFERENCE/DOCKET NUMBER: 93-15
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-547-8080 ext 322
TELEFAX: 206-548-2329
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 8878 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: human fibrinogen B-beta chain
FEATURE:
NAME/KEY: misc_RNA
LOCATION: 1..469
FEATURE:
NAME/KEY: exon
LOCATION: 470..583
FEATURE:
NAME/KEY: intron
LOCATION: 584..3257
FEATURE:
NAME/KEY: exon
LOCATION: 3258..3449
FEATURE:
NAME/KEY: intron
LOCATION: 3450..3938
FEATURE:
NAME/KEY: exon
LOCATION: 3939..4122
FEATURE:
NAME/KEY: intron
LOCATION: 4123..5042
FEATURE:
NAME/KEY: exon
LOCATION: 5043..5270
FEATURE:
NAME/KEY: intron
LOCATION: 5271..5830
FEATURE:
NAME/KEY: exon
LOCATION: 5831..5944
FEATURE:
NAME/KEY: intron
LOCATION: 5945..6632
FEATURE:
NAME/KEY: exon
LOCATION: 6633..6758

```

```

1       FEATURE:
2       NAME/KEY:      Intron
3       LOCATION:      6759..6966
4       FEATURE:
5       NAME/KEY:      exon
6       LOCATION:      6967..7252
7       FEATURE:
8       NAME/KEY:      Intron
9       LOCATION:      7253..7870
10      FEATURE:
11     NAME/KEY:      exon
12     LOCATION:      7871..8102
13     FEATURE:
14     NAME/KEY:      3'UTR
15     LOCATION:      8103..8537
16     FEATURE:
17     NAME/KEY:      misc_RNA
18     LOCATION:      8538..8878
19     FEATURE:
20     NAME/KEY:
21     LOCATION:
22     CDS
23     join(470..583, 3258..3449, 3939..4122, 5043..5270,
24     5831..5944, 6633..6758, 6967..7252, 7871..8102)
25 OS=08-206-176-5

```

	Query Match	Similarity	Score	DB 1	Length	8878;
	Best Local	Similarity	52.3%	Pred.	No. 18;	
	Matches	57;	Conservative	4;	Mismatches	48;
				Indels	0;	Gaps
QY	42	TTATCTTGAGGCGACATATTCGTATATGCGCAATTTAAATRTTCTTGCGGCATTTCC	101			
DB	3553	TTATTTTGTGTTGTTATTTTGGAAATTAATAATCAACAACATTAATGAGCCCTTTGGT	3612			
QY	102	TTACATGTCTTGACACAGTTAAATGCTGTGCCAAATTTTGATTTT	150			
DB	3613	TTAGGCTTCTTCTTGTTTCTTTGTCCTTGCGCCCAAAATTTCAAAAT	3661			

Search completed: September 25, 1999, 07:30:14
Job time: 3451 sec .

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 12:05:32 ; Search time 1811.29 seconds
(Without alignments)
417.095 Million cell updates/sec

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Title: US-09-030-606-223
Sequence: 383
1 AAAAAACAAACAAACAAAAA.....ATTACCTATATGTTTAAA 383

Scoring table: IDENTITY_NUC
Searched: 2546578 seqs, 986266752 residues

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: em_est20:*
21: em_est21:*
22: em_est22:*
23: em_est23:*
24: em_est24:*
25: em_est25:*
26: em_est26:*
27: em_est27:*
28: em_est28:*
29: em_est29:*
30: em_est30:*
31: em_est31:*
32: em_est32:*
33: em_est33:*
34: em_est34:*
35: em_est35:*
36: em_est36:*
37: em_est37:*
38: em_est38:*
39: em_est39:*
40: em_est40:*
41: em_est41:*
42: em_est42:*
43: em_est43:*
44: em_est44:*
45: em_est45:*
46: em_est46:*
47: em_est47:*
48: em_est48:*
49: em_est49:*
50: em_est50:*
51: em_est51:*
52: em_est52:*
53: em_est53:*

Result

No.

Score

Query

Match

Length

DB

ID

Description

Result No.	Score	Query Match	Length	DB	ID	Description
1	376.8	98.4	400	36	AA653016	AA653016 ns68g01.s
2	374	97.7	387	35	AA579247	AA579247 n136d10.s
3	369.8	96.6	522	26	W56075	W56075 zc56b05.r1
4	347.2	90.7	469	43	AI188401	AI188401 qd08e02.x
5	339.4	88.6	432	35	C75566	C75566 C75566 Humna
6	334.4	87.3	472	41	AI027536	AI027536 cws2c09.x
7	327.4	85.5	389	39	AA879150	AA879150 nw84e05.s
8	320.6	83.7	442	43	AI160743	AI160743 qb49d03.x
9	317.4	82.9	444	47	AI522136	AI522136 t178f03.x
10	313.4	81.8	419	35	C75584	C75584 C75584 Humna
11	303.4	79.2	486	41	AI005274	AI005274 ou07f12.x
12	291.6	76.1	473	41	AI051146	AI051146 oy49d04.x
13	288	75.2	407	36	AA622892	AA622892 np58c12.s
14	280.4	73.2	344	28	C14656	C14656 C14656 Clon
15	277	72.3	463	46	AI420653	AI420653 t13b04.x
16	261.2	68.2	324	31	AA318914	AA318914 EST21101
17	257	67.1	442	45	AI382141	AI382141 te30c04.x
18	229.6	59.9	221	36	AA653045	AA653045 ns71f07.s
19	227	59.3	268	28	AA093999	AA093999 cl1544.se
20	213	55.6	225	36	AA652561	AA652561 ns73d06.s
21	203	53.0	332	36	AA657851	AA657851 nu08d07.s
22	189.6	49.5	352	28	AA093469	AA093469 kk5000.se
23	185.4	48.4	246	36	AA641128	AA641128 nr28n07.r
24	185	48.3	372	43	AI214872	AI214872 gm31f09.x
25	181.4	47.4	200	32	AA372885	AA372885 EST84852
26	171	44.6	373	35	AA580097	AA580097 nh51h04.s
27	158.4	41.4	282	30	AA216139	AA216139 hp05z0.se
28	151.4	39.5	255	24	H83872	H83872 yv84b03.s1
29	130.8	34.2	410	21	T61899	T61899 yb86e01.s1
30	118.2	30.9	703	48	AI546846	AI546846 PZ2.1.07
31	102.8	26.8	110	34	AA503672	AA503672 ng78h10.s
32	101.2	26.4	309	21	T61960	T61960 yb86e01.r1
33	92.4	24.1	576	39	C85936	C85936 C85936 Mous
34	92.4	24.1	438	44	AI133916	AI133916 u138g06.x
35	92.4	24.1	373	46	AI450559	AI450559 mq79f07.r
36	91	23.8	425	29	AA137503	AA137503 mq79f07.r
37	83.2	21.7	231	36	AA658025	AA658025 nu15d01.s
38	76	19.8	275	49	AV003262	AV003262 AV003262
39	68.6	17.9	210	49	AV010724	AV010724 AV010724
40	64	16.7	298	50	AV044015	AV044015 AV044015
41	62.4	16.3	194	45	AI350423	AI350423 q117d03.x
42	42.6	11.1	427	46	AI152967	AI152967 ud54d06.r
43	42.6	11.1	437	42	AI428145	AI428145 ml45c02.x
44	41.2	10.8	412	44	AI289102	AI289102 gw21g01.x
45	40.8	10.7	392	23	D61957	D61957 HUM226f108

ALIGNMENTS

RESULT 1
AA653016 400 bp mRNA EST 25-NOV-1997
LOCUS AA653016
DEFINITION ns68g01.s1 NCI_CGAP_Pt2 Homo sapiens CDNA clone IMAGE:1188816, mRNA
sequence.
AA653016
ACCESSION 92584668
NID
VERSION AA653016.1 GI:2584668

Matches 371: Conservative 5; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 AAACAACAACAACAACAACAACAATCTTCAATTCAGAAAAATTAATCTAGGAGTATAT 60
    |||
Db 11 AAACAACAACAACAACAACAACAATCTTCAATTCAGAAAAATTAATCTAGGAGTATAT 70
    |||
QY 61 TGTATATATGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 120
    |||
Db 71 TGTATATATGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 130
    |||
QY 121 TTAATATGTCGTGCAAAATTTTGTATTTATTTGAGACTCTTATCAAAAGTAATGC 180
    |||
Db 131 TTAATATGTCGTGCAAAATTTTGTATTTATTTGAGACTCTTATCAAAAGTAATGC 190
    |||
QY 181 TGCACAAAGAGTCATAGGAATTAAGTATGTTCCCACTGTTTGGAGTGTCTATTC 240
    |||
Db 191 TGCACAAAGAGTCATAGGAATTAAGTATGTTCCCACTGTTTGGAGTGTCTATTC 250
    |||
QY 241 TAAAGATTTGATTTCTGGAATGACAAATTAATTAATTAATTTTAACTTTGGGGGAAAGATT 300
    |||
Db 251 TAAAGATTTGATTTCTGGAATGACAAATTAATTAATTAATTTTAACTTTGGGGGAAAGATT 310
    |||
QY 301 ATAGACACACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 360
    |||
Db 311 ATAGACACACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 370
    |||
QY 361 ACCATTAGCTATATATGT 377
    |||
Db 371 ACCATTAGCTATATATGT 387
    |||

```

```

RESULT 3
W56075 522 bp mRNA EST 11-OCT-1996
LOCUS zc56b05.f1 Soares_Parathyroid_tumor_NbHPA Homo sapiens cDNA clone
DEFINITION IMAGE:326289 5', mRNA sequence.
ACCESSION W56075
NID 91357965
VERSION W56075.1 GI:1357965
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 522)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.

```

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On Apr 14, 1993 this sequence version replaced gi:785211.

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 595 Std Error: 0.00
Seq primer: mod. REGA+ET
High quality sequence stop: 335.
Location/Qualifiers
1..522
/organism="Homo sapiens"
/db_xref="GDB:1250491"
/db_xref="taxon:9606"
/map="643H05"
/clone="IMAGE:326289"
/clone_11b="Soares_Parathyroid_tumor_NbHPA"

FEATURES
source

```

/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="Dh10B (ampicillin resistant)"
/notes="Organ: parathyroid gland; Vector: pRT3D
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer
5'-
TGTATACCAATCTGAAAGTGGAGCGCCGACCAATTTTATTTTATTTTATTTTATTTTATTTT
T-3'}, double-stranded cDNA was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pRT3
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldi. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases,
NIH."

```

BASE COUNT 171 a 71 c 94 g 181 t 5 others
ORIGIN

Query Match 96.6%; Score 369.8; DB 26; Length 522;
Best Local Similarly 97.9%; Pred. No. 8.6e-67;
Matches 368; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

```

QY 7 AACAAACAAAAAACAATCTTCATTCAGAAAAATATCTTAGGAGCTGATTTGGTAA 66
    |||
Db 1 AACAAACAAAAAACAATCTTCATTCAGAAAAATATCTTAGGAGCTGATTTGGTAA 60
    |||
QY 67 TTATGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 126
    |||
Db 61 TTATGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 120
    |||
QY 127 TGTCTGTGCCAAATTTTGTATTTATTTATTTGAGACTCTTATCAAAAGTAATGCTGCCAA 186
    |||
Db 121 TGTCTGTGCCAAATTTTGTATTTATTTATTTGAGACTCTTATCAAAAGTAATGCTGCCAA 180
    |||
QY 187 AGAAGTCTAAGGAATAGTAGTGTGTCCTCCCACTCTTTGGAGTGTGCTATTAAG 246
    |||
Db 181 AGAAGTCTAAGGAATAGTAGTGTGTCCTCCCACTCTTTGGAGTGTGCTATTAAG 240
    |||
QY 247 ATTTGATTTCTCGAATGACAAATTAATTAATTTTAACTTTGGGGGAAAGTAATAGA 306
    |||
Db 241 ATTTGATTTCTCGAATGACAAATTAATTAATTTTAACTTTGGGGGAAAGTAATAGA 300
    |||
QY 307 CCACAGCTTCACCTTCGATACCTGTAATTAATCTTTATTTGACCTGTTTGACCAT 366
    |||
Db 301 CCACAGCTTCACCTTCGATACCTGTAATTAATCTTTATTTGACCTGTTTGACCAT 360
    |||
QY 367 AAGCTATATGTTTAA 382
    |||
Db 361 AAGCTATATGTTTAA 376
    |||

```

```

RESULT 4
A1188401/c 469 bp mRNA EST 28-OCT-1998
LOCUS A1188401
DEFINITION q008e02.x1 Soares_Placenta_80yweeks_2NBHP8to9W Homo sapiens cDNA
clone IMAGE:173130 3', mRNA sequence.
ACCESSION A1188401
NID 93739610
VERSION A1188401.1 GI:3739610
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 469)
AUTHORS NC1-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

```


TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 17, 1998 this sequence version replaced gi:1899479.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bdrip/image/image.html

Insert Length: 564 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 444.
Location/Qualifiers

FEATURES

source

1. 472
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1650448"
/clone_lib="Soares_parathyroid_tumor_NbHPA"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: parathyroid gland; Vector: p773D
(Pharmacia) with a modified polylinker; Site: 1: Not I;
Site: 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer
[5'-
TGTACCAATCTGAGGAGGCGCCGACCAATTTTCTTTTCTTTTCTTTT
T-3']], double-stranded cDNA was size selected, ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified p773
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases,
NIH."

BASE COUNT 175 a 80 c 55 g 162 t
ORIGIN

Query Match 87.3%; Score 334.4; DB 41; Length 472;

Best Local Similarity 97.9%; Pred. No. 1.5e-59;
Matches 332; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

45 TCTTAGGACCTGATATGTAATATGTCATTTATATWRTTKTGGGCACTTCCTTA 104
|||||
472 TCTTAGGACCTGATATGTAATATGTCATTTATATWRTTKTGGGCACTTCCTTA 413
|||||
105 CATGTCTTGACAGATTAAATGTCGTGCCAAATTTTGTATTTATTTGGAGACTTC 164
|||||
412 CATGTCTTGACAGATTAAATGTCGTGCCAAATTTTGTATTTATTTGGAGACTTC 353
|||||
165 TTATCAAAAGTAATGTCGCCAAAGAGCTTAAGAAATTAAGTGTCCCMCACTGT 224
|||||
352 TTATCAAAAGTAATGTCGCCAAAGAGCTTAAGAAATTAAGTGTCCCMCACTGT 293
|||||
225 TTGAGAGTGTATTTTAAAGATTTTGTCTTGGAATGACATATATATTAACCTT 284
|||||
292 TTGAGAGTGTATTTTAAAGATTTTGTCTTGGAATGACATATATATTAACCTT 233
|||||
285 GGTGGGGAANAAGTTATAGACCAAGCTTCACTTCTGATACCTGTAATTAATCTT 344
|||||
232 GGTGGGGAANAAGTTATAGACCAAGCTTCACTTCTGATACCTGTAATTAATCTT 173
|||||
345 TATTGCACTGTGTTGACCAATTAAGCTATATGTTTAA 383
|||||

DB 172 TATTGCACTGTGTTGACCAATTAAGCTATATGTTTAA 134

RESULT 7
AA879150 389 bp mRNA EST 25-MAR-1998
LOCUS nw84e05.s1 NCI-CGAP_Pri12 Homo sapiens cDNA clone IMAGE:1253312,
mRNA sequence.
ACCESSION AA879150
NID 52988115
VERSION AA879150.1 GI:2968115
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 389)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2153434.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D.,
Rodrigo F. Chuquai, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bdrip/image/image.html

Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 381.
Location/Qualifiers

FEATURES

source

1. 389
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="12"
/clone_image="1253312"
/clone_lib="NCI-CGAP_Pri12"
/sex="male"
/tissue_type="metastatic prostate bone lesion"
/lab_host="DH10B"
/note="Vector: PAMP10; mRNA made from metastatic prostate
lesion of the bone. cDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 600 bp. Library made by D. Kitzman,
NIH."

BASE COUNT 121 a 51 c 75 g 142 t
ORIGIN

Query Match 85.5%; Score 327.4; DB 39; Length 389;

Best Local Similarity 95.5%; Pred. No. 4.2e-58;
Matches 342; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

26 TCTTCAATTCAGAAAATATATCTTAGGACGATATGTAATATGATGTCATTTATWRT 85
|||||
3 TCGACATTCAGAAAATATATCTTAGGACGATATGTAATATGATGTCATTTATWRT 62
|||||
86 RTTKTGGGCAATTCCTTACATGCTCTTGACAAGATTAAATGTCGTGCCAAATTTTG 145
|||||
63 ATT-TGGGCAATTCCTTACATGCTCTTGACAAGATTAAATGTCGTGCCAAATTTTG 121
|||||
146 TATTTATTTGGAGACTCTTATCAAAAGTAATGTCGCCAAAGAGCTTAAGAAATTAG 205
|||||
122 TATTTATTTGGAGACTCTTATCAAAAGTAATGTCGCCAAAGAGCTTAAGAAATTAG 181
|||||
206 TAGTGTCCCMCACTGTGTTGAGAGTGTCTATTTCAAAAGATTGATTTCTGGAATG 265
|||||

DB 182 TAGTGTCCCATCAGCTGTTGGAGTGTGCTATTCTAAAGATTTCGATTCCTGGAAG 241
 |||
 QY 266 ACAATATATTTAACTGGTGGGGAANAGTATAGACACAGCTCTCACTCTGA 325
 |||
 DB 242 ACAATATATTTAACTGGTGGGGAAGTATAGACACAGCTCTCACTCTGA 301
 |||
 QY 326 TACTGTAAATTAATCTTTATTCGACTGTTTGGACCATTAATATGTTTAA 383
 |||
 DB 302 TACTGTAAATTAATCTTTATTCGACTGTTTGGACCATTAATATGTTTAA 359
 |||

RESULT 8
 A1160743 442 bp mRNA EST 26-OCT-1998
 LOCUS gb49d03.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:1703429 3',
 DEFINITION mRNA sequence.
 ACCESSION A1160743
 NID 93694123
 VERSION A1160743.1 GI:3694123
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 442)
 NCBI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute / National Institute of Neurological
 Disorders and Stroke, Brain Tumor Genome Anatomy Project
 TITLE (CGAP/BRGAP), Tumor Gene Index
 JOURNAL Unpublished (1998)
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2286564.

CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
 Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNI at:
 www.bio.lnl.gov/db/brp/image/image.html

Insert Length: 2616 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 433.
 Location/Qualifiers

FEATURES
 source 1..442

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:1703429"
 /clone_lib="NCI_CGAP_Brn23"
 /tissue_type="g1oblastoma (pooled)"
 /lab_host="DH10B"
 /note="Organ: brain; Vector: pUT73D-Pac (Pharmacia) with a
 modified polylinker. Site 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - o1igo(dt) primer [5'
 TGTTCACATCTGAGTGGAGCGCGCATCTCTTTTCTTTTCTTTTCTTTTCTTTT
 T 3'] double-stranded cDNA was ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the pUC19 and Eco RI sites of the modified pUT73 vector.
 Library is normalized, and was constructed by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 172 a 76 c 53 g 141 t

Query Match 83.7%; Score 320.6; DB 43; Length 442;
 Best Local Similarity 97.0%; Pred. No. 1e-56;
 Matches: 320; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 54 CCGATATGCTAATATAGCTAATTAATWRTTRTGGGGCATTCCTACATGCTCT 113
 |||
 DB 442 CCGATATGCTAATATAGCTAATTAATWRTTRTGGGGCATTCCTACATGCTCT 383
 |||
 QY 114 GACAGATTAATTAATGCTGTCGCAAAATTTGATTTATTTGGAGACTCTTATCAAAA 173
 |||
 DB 382 GACAGATTAATTAATGCTGTCGCAAAATTTGATTTATTTGGAGACTCTTATCAAAA 323
 |||
 QY 174 GTAAGCTCCCAAGAGACTTAAGGAATTAAGTATGCTCCCTACTGTTGGAGGT 233
 |||
 DB 322 GTAAGCTCCCAAGAGACTTAAGGAATTAAGTATGCTCCCTACTGTTGGAGGT 263
 |||
 QY 234 GCTATCTTAAGATTTTATTCCTCGAATGACATTAATTTTACCTGGTGGGGA 293
 |||
 DB 262 GCTATCTTAAGATTTTATTCCTCGAATGACATTAATTTTACCTGGTGGGGA 203
 |||
 QY 294 AANAGTTATAGACACACAGCTCTCACTCTGATTAATTAATCTTTATGCACT 353
 |||
 DB 202 AANAGTTATAGACACACAGCTCTCACTCTGATTAATTAATCTTTATGCACT 143
 |||
 QY 354 TGTGTTGACCATTAAGCTATATGTTTAA 383
 |||
 DB 142 TGTGTTGACCATTAAGCTATATGTTTAA 113
 |||

RESULT 9
 A1522136 444 bp mRNA EST 13-APR-1999
 LOCUS t178f03.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2138141 3',
 DEFINITION mRNA sequence.
 ACCESSION A1522136
 NID 94436271
 VERSION A1522136.1 GI:4436271
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 444)
 NCBI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Mar 10, 1998 this sequence version replaced gi:2948750.

CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNI at:
 www.bio.lnl.gov/db/brp/image/image.html

Insert Length: 539 Std Error: 0.00
 Seq primer: -40UP from Gibco.
 Location/Qualifiers

FEATURES
 source 1..444

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:2138141"
 /clone_lib="NCI_CGAP_Kid11"
 /lab_host="DH10B"
 /note="Organ: kidney; Vector: pUT73D-Pac (Pharmacia) with
 a modified polylinker. Site 1: Not I; Site 2: Eco RI;
 Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
 prepared, and ss circles were made in vitro. Following HAP
 purification, this DNA was used as tracer in a subtractive
 hybridization reaction. The driver was PCR-amplified cDNAs

Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www-bio.11nl.gov/bdrrp/Image/Image.html

Contact: Robert Strausberg, Ph.D
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILM at: www-bio.11nl.gov/bdrp/image/image.html

Insert Length: 948 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 385.
 Location/Qualifiers

FEATURES

source

1..407

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1130518"

/clone_1lb="NCI-CCAP_Br2"

/sex="female, pooled"

/tissue_type="breast"

/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled bulk breast tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. This library is the normalized version of NCI-CCAP-Br1.1. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 157 a 72 c 50 g 128 t
 ORIGIN

Query Match 75.2%: Score 288; DB 36; Length 407;
 Best Local Similarity 99.0%: Pred. No. 4.7e-50;
 Matches 288; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 93 GGCATTCTCTTCATGCTGTACCAAGATTAAATGTCGTCCCAAAATTTTGATTTTA 152
 DB 407 GGCATTCTCTTCATGCTGTACCAAGATTAAATGTCGTCCCAAAATTTTGATTTTA 348
 OY 153 TTGGAGACTTCTTATCAAAAGATTAAATGTCGTCCCAAAATTTTGATTTTA 212
 DB 347 TTGGAGACTTCTTATCAAAAGATTAAATGTCGTCCCAAAATTTTGATTTTA 288
 OY 213 CCCTCAGCTTGTGGAGAGTGTCTATTCATAAAGATTGATTCCTGGAATGACAATTA 272
 DB 287 CCCTCAGCTTGTGGAGAGTGTCTATTCATAAAGATTGATTCCTGGAATGACAATTA 228
 OY 273 TATTTAACTTTGTGGGGGAAAGATTAAAGATTGATTCCTGGAATGACAATTA 332
 DB 227 TATTTAACTTTGTGGGGGAAAGATTAAAGATTGATTCCTGGAATGACAATTA 168
 OY 333 AAATTAACCTTTTATGCACTGTTTGACCACTTAAGCAATATGTTTAAA 383
 DB 167 AAATTAACCTTTTATGCACTGTTTGACCACTTAAGCAATATGTTTAAA 117

RESULT 14
 LOCUS C14656 344 bp mRNA EST 30-SEP-1996
 DEFINITION C14656 Clontech human aorta polyA+ mRNA (#6572) Homo sapiens CDNA
 accession C14656
 NID 91569363
 VERSION C14656.1 GI:1569363
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 344)

AUTHORS

Fujiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M., Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shimomiyu, H., Takaichi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y., Maekawa, H., Shin, S., and Nakamura, Y.
 Fujiwara et al. (1995)
 Unpublished (1995)
 On Sep 12, 1996 this sequence version replaced g1:1407498.

COMMENT

Contact: Tsutomu Fujiwara
 Otsuka GEN Research Institute
 Otsuka Pharmaceutical Co., Ltd
 463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
 Tel: 0886-65-2888
 Fax: 0886-37-1035.

FEATURES

Location/Qualifiers

1..344

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="GEN-078E02"

/clone_1lb="Clontech human aorta polyA+ mRNA (#6572)"

BASE COUNT 138 a 44 c 50 g 108 t

ORIGIN

Query Match 73.2%: Score 280.4; DB 28; Length 344;
 Best Local Similarity 96.5%: Pred. No. 1.7e-48;
 Matches 276; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAACCAACCAACCAACCAACCAATTCCTTCATTCAGAAAATTAATTCAGGAGATAT 60
 DB 59 AAACCAACCAACCAACCAACCAATTCCTTCATTCAGAAAATTAATTCAGGAGATAT 118
 OY 61 TGTATATTATGTCATTAATTAATTTTGTGGGCGATTCCTTACATTTGCTTGAACA 120
 DB 119 TGTATATTATGTCATTAATTAATTTTGTGGGCGATTCCTTACATTTGCTTGAACA 178
 OY 121 TTAATATGTCGTCCCAAAATTTGATTTATTTGAGACTTCTTACCAAGATTAAGC 180
 DB 179 TTAATATGTCGTCCCAAAATTTGATTTATTTGAGACTTCTTACCAAGATTAAGC 238
 OY 181 TGCCAAAGAGTCTAAGAAATTAAGTATGTTCCCTCAGCTTGTGGAGTGTCTATTC 240
 DB 239 TGCCAAAGAGTCTAAGAAATTAAGTATGTTCCCTCAGCTTGTGGAGTGTCTATTC 298
 OY 241 TAAAGATTGATTTCTGGAATGACATTAATTTTAACTTTGG 286
 DB 299 TAAAGATTGATTTCTGGAATGACATTAATTTTAACTTTGG 344

RESULT 15
 LOCUS A1420653/c 463 bp mRNA EST 28-MAR-1999
 DEFINITION A1420653 t13b04.x1 NCI-CCAP_Brn23 Homo sapiens CDNA clone IMAGE:2096047 3', mRNA sequence.
 accession A1420653
 NID 9426584
 VERSION A1420653.1 GI:4266584
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 463)
 AUTHORS NCI/NINDS-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CCAP/BRGAP), Tumor Gene Index
 JOURNAL Unpublished (1998)
 COMMENT On Mar 20, 1998 this sequence version replaced g1:2980399.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
www-bio.lnl.gov/brp/image/image.html

Insert Length: 1420 Std Error: 0.00

Seq. primer: -400P from Gibco

High quality sequence stop: 449.

Location/Qualifiers

1. 463

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image="IMAGE:2096047"

/clone_lib="NCI-CGAP_Brn23"

/issue_type="glioblastoma (pooled)"

/lab_host="DH10B"

/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dt) primer (5'
TGTACCAATCTGAGAGGAGCGCGCGCATATCTTTTCTTTTCTTTTCTTTT
T 3'); double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 179 a 76 c 55 g 153 t
ORIGIN

Query Match 72.38; Score 277; DB 46; Length 463;
Best Local Similarity 98.68; Pred. No. 8.1e-48;
Matches 288; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

```

QY 92 GGGCATTTCCTTACATGTCCTGACAGATTAATGTCGTGCCAAATTTGTAATTT 151
   |||||||
Db 463 GGGCATTTCCTTACATGTCCTGACAGATTAATGTCGTGCCAAATTTGTAATTT 404

QY 152 ATTTGAGACTTCTTATCAAAAGTAATGCTGCCAAGGAAGCTTAAGAAATTAGTAGTGT 211
   |||||||
Db 403 ATTTGAGACTTCTTATCAAAAGTAATGCTGCCAAGGAAGCTTAAGAAATTAGTAGTGT 345

QY 212 TCCGTCACCTGTTGGAGTGTGCTATTTCTAAAGATTTTCTGCGAATGACAATT 271
   |||||||
Db 344 TCCGTCACCTGTTGGAGTGTGCTATTTCTAAAGATTTTCTGCGAATGACAATT 285

QY 272 ATATTTTAACTTGGTGGGGAANAAGTATAGGACCAAGCTTCACTTCTGATCTG 331
   |||||||
Db 284 ATATTTTAACTTGGTGGGGAANAAGTATAGGACCAAGCTTCACTTCTGATCTG 225

QY 332 TAAATTAATCTTTTATGCACTGTTTGAACATTAAAGCTATATGTTTAAA 383
   |||||||
Db 224 TAAATTAATCTTTTATGCACTGTTTGAACATTAAAGCTATATGTTTAAA 173

```

Search completed: September 25, 1999, 12:05:37
Job time: 8141 sec

PR 23-MAY-1997; US-047597.
 PR 23-MAY-1997; US-047598.
 PR 23-MAY-1997; US-047599.
 PR 23-MAY-1997; US-047600.
 PR 23-MAY-1997; US-047601.
 PR 23-MAY-1997; US-047612.
 PR 23-MAY-1997; US-047613.
 PR 23-MAY-1997; US-047614.
 PR 23-MAY-1997; US-047615.
 PR 23-MAY-1997; US-047617.
 PR 23-MAY-1997; US-047618.
 PR 23-MAY-1997; US-047632.
 PR 23-MAY-1997; US-047633.
 PR 06-JUN-1997; US-048964.
 PR 06-JUN-1997; US-048974.
 PR 22-AUG-1997; US-056630.
 PR 22-AUG-1997; US-056631.
 PR 22-AUG-1997; US-056632.
 PR 22-AUG-1997; US-056636.
 PR 22-AUG-1997; US-056637.
 PR 22-AUG-1997; US-056662.
 PR 22-AUG-1997; US-056664.
 PR 22-AUG-1997; US-056845.
 PR 22-AUG-1997; US-056852.
 PR 22-AUG-1997; US-056864.
 PR 22-AUG-1997; US-056872.
 PR 22-AUG-1997; US-056874.
 PR 22-AUG-1997; US-056875.
 PR 22-AUG-1997; US-056876.
 PR 22-AUG-1997; US-056877.
 PR 22-AUG-1997; US-056878.
 PR 22-AUG-1997; US-056879.
 PR 22-AUG-1997; US-056880.
 PR 22-AUG-1997; US-056881.
 PR 22-AUG-1997; US-056882.
 PR 22-AUG-1997; US-056884.
 PR 22-AUG-1997; US-056886.
 PR 22-AUG-1997; US-056887.
 PR 22-AUG-1997; US-056888.
 PR 22-AUG-1997; US-056889.
 PR 22-AUG-1997; US-056893.
 PR 22-AUG-1997; US-056894.
 PR 22-AUG-1997; US-056903.
 PR 22-AUG-1997; US-056908.
 PR 22-AUG-1997; US-056909.
 PR 22-AUG-1997; US-056910.
 PR 22-AUG-1997; US-056911.
 PR 05-SEP-1997; US-057650.
 PR 05-SEP-1997; US-057761.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
 PI Feng P, Fertle AM, Fischer CL, Graves KA, Greene DM, Hu JS,
 PI Kyau H, Lafleur DM, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
 DR WPI: 98-609887/51.
 N-PEDB: V34154.
 CC New isolated human genes and the secreted polypeptides they encode
 CC useful for diagnosis and treatment of e.g. cancers, neurological
 CC disorders, immune diseases, inflammation or blood disorders
 CC Claim 1; Page 278-280; 447pp; English.
 CC This sequence represents a secreted human protein encoded by the gene
 CC clone detailed in the descriptor line.
 CC The gene can be used to generate fusion proteins by linking to the gene
 CC to a human immunoglobulin Fc portion (e.g. V34145) for increasing the
 CC stability of the fused protein as compared to the human protein only.
 CC The invention relates to 70 novel genes and their fragments (nucleic acid
 CC sequences: V34154-V34276; amino acid sequences W75057-W75179) which
 CC are useful for preventing, treating or ameliorating medical conditions
 CC e.g. by protein or gene therapy. Also, pathological conditions can be
 CC diagnosed by determining the amount of the new polypeptides in a sample
 CC or by determining the presence of mutations in the new polynucleotides.
 CC *Specific* uses are described for each of the 70 polynucleotides, based on

CC which tissues they are most highly expressed in (see V34154 for described
 CC uses).
 CC Sequence 467 AA;
 alignment_scores:
 Quality: 71.00 Length: 75
 Ratio: 1.919 Gaps: 3
 Percent Similarity: 49.333 Percent Identity: 30.667
 alignment_block:
 US-09-030-606-224/rev x W75057 ..
 Align seg 1/1 to: W75057 from: 1 to: 467
 229 TATCAAGACACCTTTAAGGCTTCAT..... 203
 180 TYTYTSerLeuPheGluTrpPheHisProLeuPheGluAspIuSe 196
 202TTYGAGAAAGTAGTATT.....TCCTTCCTTC 172
 196 rSerSerPheHisLysArgGlnPheProValSerLysTrpLeuProGlu 213
 213 eutyGluLeuValAsnAsnTyGlnProGluValLeu..... 225
 121 CATTTTTTTGTGATGGAGGATACAGGCGTACACTCCCTACTACATGT 72
 226TrpSerAspLysAspLysAlaProAspGlnTyTrpAsn** 240
 71 CACAACCTTTCTTTTGTGTC 47
 240 *thrgLypheLeuAlaTrpLeuTy 248
 seq_name: A_Geneseq_36:Y13378
 seq_documentation_block:
 ID Y13378 standard; Protein; 467 AA.
 AC Y13378:1999 (first entry)
 DT 25-JUN-1999
 DE Amino acid sequence of protein PRO260.
 KW Secreted protein; transmembrane protein; human; enterocolitis;
 KW Zollinger-Ellison syndrome; gastrointestinal ulceration;
 KW congenital microvillus atrophy; skin disease; cell growth;
 KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
 KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
 KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;
 KW anti-thrombotic; wound healing; tissue repair.
 OS Homo sapiens.
 PN M09914328-A2.
 PD 25-MAR-1998.
 PF 16-SEP-1998; U19330.
 PR 25-NOV-1997; US-066840.
 PR 17-SEP-1997; US-059113.
 PR 17-SEP-1997; US-059115.
 PR 17-SEP-1997; US-059117.
 PR 17-SEP-1997; US-059119.
 PR 17-SEP-1997; US-059121.
 PR 17-SEP-1997; US-059122.
 PR 17-SEP-1997; US-059184.
 PR 18-SEP-1997; US-059263.
 PR 18-SEP-1997; US-059266.
 PR 15-OCT-1997; US-062125.
 PR 17-OCT-1997; US-062285.
 PR 17-OCT-1997; US-062287.
 PR 21-OCT-1997; US-063486.
 PR 24-OCT-1997; US-062814.
 PR 24-OCT-1997; US-062816.
 PR 24-OCT-1997; US-063045.
 PR 24-OCT-1997; US-063120.
 PR 24-OCT-1997; US-063121.
 PR 24-OCT-1997; US-063127.

PR 24-OCT-1997: US-063128.
PR 27-OCT-1997: US-063329.
PR 27-OCT-1997: US-063327.
PR 28-OCT-1997: US-063541.
PR 28-OCT-1997: US-063542.
PR 28-OCT-1997: US-063544.
PR 28-OCT-1997: US-063549.
PR 28-OCT-1997: US-063550.
PR 28-OCT-1997: US-063564.
PR 29-OCT-1997: US-063435.
PR 29-OCT-1997: US-063704.
PR 29-OCT-1997: US-063732.
PR 29-OCT-1997: US-063738.
PR 29-OCT-1997: US-063734.
PR 29-OCT-1997: US-064215.
PR 31-OCT-1997: US-063735.
PR 31-OCT-1997: US-063870.
PR 31-OCT-1997: US-064103.
PR 03-NOV-1997: US-064248.
PR 07-NOV-1997: US-064809.
PR 12-NOV-1997: US-065186.
PR 17-NOV-1997: US-065846.
PR 18-NOV-1997: US-065693.
PR 21-NOV-1997: US-066120.
PR 21-NOV-1997: US-066364.
PR 24-NOV-1997: US-066772.
PR 24-NOV-1997: US-066466.
PR 24-NOV-1997: US-066770.
PR 24-NOV-1997: US-066511.
PR 24-NOV-1997: US-066453.
PA (GETH) GENENTECH INC.
PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
PI WPI; 99-229533/19.
PR N-SDB: X52249.
PT New isolated human genes and polypeptides used in, e.g. treatment of
PT gastrointestinal ulceration
PS CLAIM 12, Fig 72; 320pp; English.
PS Y13344-403 represent secreted and transmembrane human proteins.
CC The cDNA sequences are obtained from cDNA libraries, prepared from
CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
CC The encoded polypeptides have specific uses based on their homology to
CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
CC associated with the preservation and maintenance of gastrointestinal
CC mucosa and the repair of acute and chronic mucosal lesions
CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
CC ulceration and congenital microvillus atrophy), skin diseases associated
CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
CC potent effects on cell growth and development, diseases related to growth
CC or survival of nerve cells including Parkinson's disease, Alzheimer's
CC disease, ALS, neuropathies or cancer. PRO265 can be used as for
CC fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
CC of Usher Syndrome or Atrophia areata; PRO269 can be used as an
CC anti-thrombotic agent; PRO287 polypeptides and portions may have
CC therapeutic applications in wound healing and tissue repair; PRO317 can
CC be used for treating problems of the kidney, uterus, endometrium, blood
CC vessels, or related tissue, e.g. in the heart of genital tract.
SQ Sequence 467 Aa;

```
alignment_scores:
  Quality: 70.00
  Ratio: 1.892
  Percent Similarity: 49.333
  Length: 75
  Gaps: 3
  Percent Identity: 30.667
```

US-09-030-606-224/rev x Y13378

Align seg 1/1 to: Y13378 from: 1 to: 467

229 TATCAAAGCACCTTTAAGGCTTCAT..... 203
 ||| ||| |||:: |||||

[illegible]

seq_name: A_Geneseq_36:R88732

seq_documentation_block:

AC R88732;

S. aureus topoisomerase IV

KW DNA gyrase; PCR; amplification; primer;

KW termination codon; Initiation; frameshift; translation; fluoroquinolone; inhibition; translation; detection

OS Staphylococcus aureus.
PN W08603515-21

PD 08-FEB-1996.
DE 26-MR-1995.

PR 27-JUL-1994; FR-009

PI Blanche F, Cameron B, Crouzet J, Famechon A, Ferrero L;
WPI: 06-117050/13

DR N-PSDB; T12563.

PT useful e.g. to screen for potential

CC This is the amino acid sequence of the Gr1B subunit of the Staphylococcus

CC other extensively studied member of the family is E.coli DNA gyrase.

CC a protein which has a pI of 7.21 and mol. wt. ca. 74.3 kD. The corresp.

CC sequence similarity with other bacterial topoisomerase IV genes e.g. the *E. coli* and *B. subtilis* *CuvB* or *E. coli* *DnaB* genes. The 1 kb fragment

CC amplified, corresp. to residues 98-430 of E.coli GyrB, was used to probe

CC the operon, the *griB* gene termination codon straddles the *griA* initiation

CC accurately translate the *gryA* mRNA. A similar linkage is seen in the
CC *gryA* and *B* genes in *Halobacterium* sp. *Moitaka* (1994) for the

CC fluoroquinolone antibacterials. The protein can thus be used to screen for new antibacterials and to screen for and inhibit ATP-dependent

CC DNA relaxation and DNA decatenation.
50 Sources: 553 11.

SQ Sequence 663 AA;

```
alignment_scores:..
    Quality:      66.00      Length:      72
    Ratio:        1.650      Gaps:        3
Percent Similarity: 55.556  Percent Identity: 30.556
```

alignment_block:

US-09-030-606-224 x R88732

Align seg 1/1 to: R88732 from: 1 to: 663

43 ATAGGAACAACAAAAAGTTGTGACATTGTAGTAGGAGTGTGT 92

```

|||||.....|||
29 IleglyserThrasplysargglyleuHshlsleuValtygluileu 45
93 ACCCTTACTCCCATCAAAAAAATGATACATGTTAAAGATARA 142
45 IAspanserValaspGluValleuasnGLyTYr..... 56
143 AGGCAATATTTATCATATGTTCTAAAGAGAGAGAAATACGA 192
57 ..GlyAsnGluileaspValthritile...AsnlyAspelyserIleSer 71
193 CTTTCTCAAAATGAAGCCCTTAAAGTCTTGATGACTGAAGACAA 242
72 IlegluAspasnGLy.....ArgglyMetProthnglyIleHshly 85
243 ATGTGGCCGTCCATCC 258
85 sSerGlylyspProthr 90
seq_name: A_Geneseq_36:Y12066

```

```

seq_documentation_block:
ID Y12066 standard; Protein: 100 AA.
AC Y12066;
DE 18-JUN-1999 (first entry)
DE Human 5' EST secreted protein SEQ ID NO: 379.
KW Human; secreted protein; EST: expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haemostasis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemokine; chemokine; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.
OS Homo sapiens.
PN M09906554-A2.
PD 11-FEB-1999.
PE 31-JUL-1998; IB1238.
PF 01-AUG-1997; US-905134.
PA (GIST ) GENSET.
PI Ducleert A, Dumas Mline Edwards J, Lacroix B;
DR NPI: 99-153784/13.
DR N-PSDB; X40899.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from kidney, fetal kidney, dystrophic
PT muscle, muscle and heart tissue
PS Claim 34: Page 500: 622pp; English.
CC X40826 to X41093 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y01602 and
CC Y11994 to Y12260, respectively. The proteins given represent the signal
CC peptide and an N-terminal fragment of a secreted protein. The nucleic
CC acid sequences can be used for producing secreted human gene products.
CC They can also be used to develop products for diagnosis and therapy.
CC The proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokine activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used
CC for directing extracellular secretion of a polypeptide or the insertion
CC of a polypeptide into a membrane, or importing a polypeptide into
CC a cell.
SQ Sequence 100 AA;

```

```

alignment_scores:
Quality: 63.50      Length: 63
Ratio: 2.352       Gaps: 3
Percent Similarity: 42.857   Percent Identity: 26.984

alignment_block:
US-09-030-606-224/rev x Y12066

```

```

Align seg 1/1 to: Y12066 from: 1 to: 100
135 TTTAACCATGATCATCTTTTCTTTGATGGGAGTAGGCG..... 94
11 PheserHsAlasSerleuPhePheMetValGlySerThrGlySerleu 27
93 .....TACACATCCCTACTACA..... 76
27 eleuLeuThrSerCysPheThrIleValSerSerThrPheLeuGln 44
75 .....ATG 73
44 yslserSerleuLeuLeuIleuPheThrGluThrSerValleuMet 60
72 TCACAACTTT.....TCTTTTGTTGTTCCATTGG 40
61 LeuLySerThrPheValAlaAsnSerCysCys***LeuTrp 73
seq_name: A_Geneseq_36:R77168

```

```

seq_documentation_block:
ID R77168 standard; Protein: 202 AA.
AC R77168;
DE 04-DEC-1995 (first entry)
DE Arabidopsis condensensing enzyme clone CE17.
KW Arabidopsis; condensensing enzyme.
OS Arabidopsis thaliana.
PN W09515387-A.
PD 08-JUN-1995.
PE 30-NOV-1994; U13686.
PF 30-NOV-1993; US-160602.
PR 23-JUN-1994; US-265047.
PA (CALY ) CALSINE INC.
PI Lardizabal KD, Lassner MW, Metz JG;
DR N-PSDB; Q90214.
PT Production of very long chain fatty acid(s) in plant(s) - to produce
PT drought and stress resistant transgenic plant(s)
PS Example; Figure 9; 148pp; English.
CC The CE15 and CE20 Brassica CDNA sequences (see Q90210,Q90211
CC and Q90212) and the condensensing enzyme encoding sequence from
CC Jojoba (Q90208) were used in determining primers Q90221-
CC Q90225 from conserved AAs. These primers were variously used to
CC PCR (RT-PCR) amplify fragments from RNA isolated from developing
CC seeds of Lunaria annua, Tropaeolum majus (Nasturtium), and
CC green illiagues of Arabidopsis thaliana. The primers most
CC successfully utilised were Q90221 and Q90222. These primers were
CC used to produce three clones encoding a portion of the elongase
CC condensensing enzyme from Arabidopsis, designated ARAB CE15, ARAB CE17
CC and ARAB CE19 (see Q90213/R77167; Q90214/R77168; Q90215/R77169
CC respectively. Clones were also identified from Lunaria and
CC Nasturtium.
SQ Sequence 202 AA;

```

```

alignment_scores:
Quality: 63.50      Length: 68
Ratio: 1.477       Gaps: 5
Percent Similarity: 63.235   Percent Identity: 33.824

```

```

alignment_block:
US-09-030-606-224/rev x R77168

```

```

Align seg 1/1 to: R77168 from: 1 to: 202
205 CATTYTGAGAAAGTAGTATTTCTCTCTCTCTCTTGAACATANGAT 156
12 HsPhePhePhePhePhePhePhe...LeuProleuMetAlaValleuPhe 27
155 AAAAAATATGCGCTTATCATCAATGATGATCAATTTTGTGATG 106
27 t.AsnValSerleuLeuSerleuAsnHsleuGln..... 38

```



```

105 GGGAGTAAAGGGGACACACCTCCTACTACAATGTCAACAACTTTCCTT 56
39 .....Leu..TyrTyrAsnSerThrGlyPheIlePhe 48
55 TTGTGTGTCCTATTGGTTGA...ACTGACATATTTTCTAACAGAGGC 9
   : : : : : | | | : : : : : | | : : : : : |
49 ValIleHrLeuAlaIleValGlySerIleValAlaPheMetSerArgPr 65
      8 T 8
      |
65 O 65

```

seq_name: A_Geneseq_36:Y07052

```
seq_documentation_block:
ID Y07052 standard; Protein; 312 AA
"07052"
```

DT 02-JUL-1999 (first entry)
DE breast cancer associated antigen precursor sequence.
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
OS Homo sapiens.
PN WO9904265-A2.
PD 28-JAN-1999.
PF 15-JUL-1998.
PR 22-JUN-1998; US-1023322.
PR 17-JUL-1997; US-896164.
PR 10-OCT-1997; US-061599.
PR 10-OCT-1997; US-061765.
PR 10-OCT-1997; US-948705.
PR 11-OCT-1997; GB-021697.
PA (LDDM-) LDDMIG INST CANCER RES.
PI Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ,
PI Pfeifferschuh M, Sahin U, Scanlan MD, Stockert E,
PI Tureci O;
DR MPI; 99-132448/11.
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
PS Disclosure; Page 439-440; 787pp; English.
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
SQ Sequence 312 AA;

```
alignment_scores:
  Quality: 62.50
  Ratio: 2.155
  Gaps: 3
Percent Similarity: 46.032
Percent Identity: 34.921
```

alignment_block:

Align seg 1/1 to: Y07052 from: 1 to: 312

115 AAAATGATACATCGTTAAAGATARAAGGC..... 147
 |||:::|||||:::|||||
 186 LysGInGLyTyrMetMetLys.....LysGLyHisArgGLyAsnTr 200

```

148 .....AATTTTAAACAATGTT... 165
200 pthmglaatgpphevalleuylsproinslelleserityryvalS 217
166 .....CTAAAGAGAGAGAGAGAAATACTACTTCTCRAAATGGA 207
      |||||:::||||| |||||::: ||
217 ergluaspheulysaspyrlysglyaspilieleuinspeliuasnucys 233
208 AGCCCTTAAAGGTGCTTANTACTGAAGACACAATGT 246
      |||:::|||| |||||
234 CysValaGluSerLeuProaspIysaspelIylsystcys 246

```

seq_name: A_Geneseq_36:W97757

```
seq_documentation_block:  
ID      W97757 standard; Protein; 585 AA  
CD      NOTSET
```

DT	21-MAY-1999	(first entry)	
DE	S-region transfer activity-70 (SRPA-70).		
KW	SRPA-70; S-region transfer activity-70; recombination protein;		
KW	switch recombination; mouse; immunomodulator; cancer; allergy;		
KW	therapy.		
OS	Mus sp.		
FH	Key	Location/Qualifiers	
FT	Peptide	222..225	
FT		/note="nuclear localisation signal"	
FT	Peptide	291..294	
FT		/note="nuclear localisation signal"	
FT	Peptide	360..376	
FT		/note="nuclear localisation signal"	
PN	MO9903991-A1.		
PD	28-JAN-1999.		
PE	16-JUL-1998.	IB1191.	
PR	18-JUL-1997.	EP-112326.	
PA	(HOEF) HOFFMANN IA ROCHE & CO AG F.		
PA	(REGC) UNIV CALIFORNIA.		
PI	Jessberger R, Mabl M;		
PI	WPI: 99-132335/11.		
DR	N-PSDB: X07259.		
PT	New DNA encoding S-region transfer activity-70 protein - used to		
PT	identify specific binding agents and for modulating the immune		
PT	response e.g. in cases of cancer and allergy		
PS	Claim 8; Fig 4; 37pp; English.		
CC	This polypeptide comprises murine SRPA-70, or S-region transfer		
CC	activity-70, a 70 kDa protein that mediates switch recombination.		
CC	The amino acid sequence was deduced from a cDNA clone (see X07259)		
CC	isolated from a mouse spleen library. The invention relates to the		
CC	isolation, purification and characterisation of proteins mediating		
CC	switch recombination. It also relates to recombinant SRPA-70		
CC	proteins, vectors and host cells. SRPA-70 polypeptides are used:		
CC	(i) for modulating the immune response; (ii) for identifying		
CC	specific (antagonists; (iii) to identify proteins (and their		
CC	genes) involved in class-switch recombination or other B cell		
CC	functions, and (iv) to raise antibodies.		
Sequence	585 AA;		

```
alignment_scores:          Length:    63  
      Quality:         62.50  
        Ratio:         2.155       Gaps:     3  
Percent Similarity:   46.032   Percent Identity: 34.921
```

alignment_block:

Align seg 1/1 to: W97757 from: 1 to: 585

115 AAAATGATACACTGGTTAAAGATARAAGGC..... 147
 |||:::|||||::||| |||||
 213 LysGlnGlyTyrMetMetLys....LysGlyHisLysArgLysAsnTr 227
 148AATATTATCATATGTT... 165

227 phtgltartrpPheValLeuysProAsnIleIleSerTyrValS 244
ID W88448 standard; Protein: 1296 AA.
166CTAAAGAGAGAGAGAAATACATTCTCRAAAGGA 207
244 ercluaspluLysAspLysLysLysLysLysLysLysLysLys 260
208 AGCCCTTAAAGGCTTGTGATCTAGAGACACAAATGT 246
261 CysValGluSerLeuProAspLysAspGlyLysAspCys 273

seq_name: A_Geneseq_36:R20030

seq_documentation_block:

ID R20030 standard; Protein: 633 AA.
AC R20030;
DT 07-APR-1992 (first entry)
DE Bacillus thuringiensis CryIIb crystal toxin.
KM B.t. crystal protein; insecticide; Lepidoptera; larva; corn earworm;
KW Heliothis zea.
OS Bacillus thuringiensis.
PN US5073632-A.
PD 17-DEC-1991.
PR 11-JUL-1989; 379015.
PR 16-APR-1987; US-039542.
PR 11-JUL-1989; US-379015.
PA (ECOG-) ECOGEN INC.
PI Donovan WP.
DR WPI: 92-016224/02.
DR N-PSDB: Q20202.
PT Bacillus thuringiensis cryIIb protein gene - used to obtain the
PT protein for use as an insecticidal cpd. against lepidopteran
PT insects
PS Disclosure: Fig 6; 30pp; English.
CC The cryIIb gene was isolated from a B. t. strain using the cryIIa
CC gene as probe. The cryIIb gene encodes a 633 amino acid protein of
CC calculated mol. wt. 70,749. The insecticidal toxin cryIIb is twice
CC as toxic as the CryIIa protein against the lepidopteran Heliothis
CC zea.
SQ Sequence 633 AA:

alignment_scores:
Quality: 59.00 Length: 81
Ratio: 1.311 Gaps: 5
Percent Similarity: 55.556 Percent Identity: 35.802

alignment_block:

US-09-030-606-224/rev x R20030 ..

Align seg 1/1 to: R20030 from: 1 to: 633

241 TGGTCTTATGATCAAGCAGCTTAAAGGCTTCATTTGAGAAAGT 192
223 CysIleAsnThrTyrGlnSerAlaPheLysGlyLeuAsnThrArg.... 237
191 AGATATTTCTCTCTCTCTCTTATGACATATGATTAATATTCCTT 142
238LeuHisAspMetLeuGlnPheArgTln.TyrMetPheLeu 230
141 YATATCTTAAACATGATCATTT.....TTTTT.....TT 110
251 AsnValPheGluTyrValSerIleTyrSerPhePheLysTyrGlnSer 267
109 GATGGGAGTAAGGGATACACTCCCTACTACAAATGTCA.....C 69
267 uLeuValIleSerSerGlyAlaAsnLeuTyrAlaSerGlySerGlyProGln 284
68 AAATTTTCTTTTGTGTTCTCTATGAGTTGTAACGTAC 28
284 InThrGlnSerPheThrSerGlnAspTyrProPheLeuTyr 297

seq_name: A_Geneseq_36:W88448

seq_documentation_block:
ID W88448 standard; Protein: 1296 AA.

AC W88448;
DT 26-APR-1999 (first entry)
DE Caenorhabditis elegans NPc1 protein orthologue.
KM Niemann-Pick disease type C; NPc1 gene; worm; orthologue;
KW diagnosis; therapy; animal model; cholesterol; neurodegeneration.
OS Caenorhabditis elegans.
PN W09901555-A1.
PD 14-JAN-1999.
PR 02-JUL-1998; U13862.
PR 03-JUL-1997; US-051682.
PA (USSH) US DEPT HEALTH & HUMAN RESOURCES.
PI Carstea ED, Gu J, Loftus SK, Morris JA, Pavan WJ,
PI Penterchev PG, Rosenfeld MA, Tagle DA;
DR WPI: 99-106056/09.
DR N-PSDB: X06876-77.
PT New isolated gene, NPC-1 - is associated with Niemann-Pick type C
PT disease, used to develop products for the study, diagnosis and
PT therapy of the disease
PS Disclosure: Page 91-95; 101pp; English.
CC This polypeptide comprises the Caenorhabditis elegans orthologue
CC of the human NPc1 polypeptide that is associated with
CC Niemann-Pick disease type 2 (NP-C). The polypeptide shows
CC extensive identity (30%) and similarity (55%) to the human
CC NPc1 protein (see W88445). Biochemical and genetic analysis of
CC yeast, worm and murine NPc1 model systems will provide resources
CC for understanding the role of NPc1 in intracellular cholesterol
CC homeostasis and in the aetiology of neurodegeneration in NP-C
CC disease. The provision of the human NPc1 cDNA sequence (see
CC X06873) enables methods of detecting the presence of mutations in
CC the NPc1 gene, and thereby facilitates the determination of
CC whether an individual is an NP-C sufferer or carrier.
SQ Sequence 1296 AA:

alignment_scores:
Quality: 59.50 Length: 114
Ratio: 1.062 Gaps: 7
Percent Similarity: 49.123 Percent Identity: 22.807

alignment_block:

US-09-030-606-224/rev x W88448 ..

Align seg 1/1 to: W88448 from: 1 to: 1296

278 AGTCATGCACCTTTAAAGAGATGGA.....CGGCCA..... 246
1153 SerHisIleSerAsnValGluGluGlyIleLeuAsnArgProSerLeu 1169
245CATTTGTCCTTCATATCAAGCAGCTTAAAGGCTTC 206
1169 uAspAlaSerHisIleLeuAspProLeuLeuLysAlaGluGlyIle. 1185
205 CATTTGAGAAAGTATTTCTCTCTCTCTTGAACATATGAT 156
1186AspLysAlaIleGlyAlaGspPheLysTyrHisAsnArg 1199
155 AAAATATGGCCTTATACCTTA.....ACCATGAT..... 123
1200 SerLeuLeuProIleTyrAlaIlePheThrAlaMetTyrLeuProAsn 1216
122CATTTT..... 114
1216 aLysSerSerTyrArgThrGlySerProIlePheValAspThrAspLeu 1233
113 ..TTTGAAGGGAGTAAGGGATACAC.....TCCCA 81
1233 euPheThrTyrSerIleProTyrIleHisPheValIleSerValSerPhe 1249
80 CTACAATGTCACAACTTTTCTTTTGTGTTGTTCCATTTGTT 39
1250 IleLysCysHisValLeuTyrIleTyrAlaIleThrIleGly 1263

seq_name: A_Geneseq_36:W39159

seq_documentation_block:

ID W39159 standard; Protein; 196 AA.
AC W39159;
DT 27-APR-1998 (first entry)
DE Clone PRBS3FH2910 #4.1 CFH related protein fragment.
KW Complement factor H; tumour associated antigen; renal cancer;
KW urogenital cancer; medicament; modulator.
OS Synthetic.
PN WO9738136-A1.
PD 16-OCT-1997.
PF 09-APR-1997; US-5710.
PR 06-MAR-1997; US-812481.
PR 09-APR-1996; US-015083.
PR 09-APR-1996; US-630048.
PR 06-MAR-1997; US-038614.
PA (BARD-) BARD DIAGNOSTIC SCI INC.
PI Enfield DL, Hase GM, Klanders RJ;
PI WPI: 97-512742/47.
DR N-PSDB; V02795.
PT Treating or screening for cancer, e.g. renal or urogenital cancer -
PT by modulating or detecting tumour associated human complement factor
PS Example 6B; Fig 7B; 104pp; English.
CC This partial protein is found in clone PRBS3FH2910 #4.1 and represents a
CC complement factor H related protein with homology to a region of the
CC human tumour-associated complement factor H (CFH). The detection of this
CC protein and a CFH antigen can be used in screening or for the treatment
CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.
CC Agents that may modulate this antigen could be used in the manufacture of
CC a medicament for the treatment of a tumour cell.
SQ Sequence 196 AA;

alignment_scores:

Quality: 57.50 Length: 90
Ratio: 1.223 Gaps: 5
Percent Similarity: 52.222 Percent Identity: 27.778

alignment_block:

US-09-030-606-224 x W39159 ..

Align seg 1/1 to: W39159 from: 1 to: 196

```
82 AGGAGTGTGTACCCCTTACTCCCATCAAAAAATGATACATGCT 131
|||||
53 ArgAspThrSerCysValAsnProProThrValGlnAlaTyrIleVal 69
132 T.....AAAGATABAAGGCAATTTTATCATATGCTTC 166
69 IserArgGlnMetSerLysTyrProSerLysGlnValArgValArgTyrGlnC 86
167 TAAAGAG.....AAGAGAGAGAAATACTACTTCT 198
86 ysArgSerProTyrGlnMetPheGlyAspGlnIValMetCysLeuAsn 102
199 CRAAATGGAAC...CCTTAAAGTGCCTTGTACTGAGACACAAATG 245
103 GlyAsnTPThrGlnProProGlnCysLysAspSerThrGly...LysC 118
246 TGCCGCTCATCTCTTATAGTTCAGTTCAGTTCAGACAGGTAACGT 295
118 sglYProPro.ProProIleAspAsnGlyAspIleThrSerPheProleu 134
296 GCAGTTTAACTACGCA 313
135 SerValTyrAlaProAla 140
```

seq_name: A_Geneseq_36:Y07068

seq_documentation_block:

ID Y07068 standard; Protein; 412 AA.

AC Y07068;

DT 02-JUL-1999 (first entry)
DE Renal cancer associated antigen precursor sequence.
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; lung cancer;
KW prostate cancer.
OS Homo sapiens.
PN WO9904265-A2.
PD 28-JAN-1999.
PF 15-JUL-1998; U14679.
PR 22-JUN-1998; US-102322.
PR 17-JUL-1997; US-896164.
PR 10-OCT-1997; US-061599.
PR 10-OCT-1997; US-061766.
PR 10-OCT-1997; US-948705.
PR 11-OCT-1997; GB-021697.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ,
PI Pfeundachun M, Sahin U, Scanlan MJ, Stockert E,
PI Threlk O,
DR WPI: 99-132448/11.
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
PS Disclosure: Page 469-470; 787pp; English.
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
SQ Sequence 412 AA;

alignment_scores:

Quality: 58.00 Length: 62
Ratio: 1.706 Gaps: 2
Percent Similarity: 54.839 Percent Identity: 27.419

alignment_block:

US-09-030-606-224 x Y07068 ..

Align seg 1/1 to: Y07068 from: 1 to: 412

```
82 AGGAGTGTGTACCCCTTACTCCCATCA.....AAAAAA 119
|||||
136 ArgAspCysArgLeuSerThrGlnHisGlnIleuSerAlaAspGlnIle 152
120 TGATACATGTTAAAGATTAAGGCAATTTTATCATATGTTCTAA 169
152 sglYSerClnIleGlnIleuLysGlyAspIleIleSerValIleIleA 169
170 AAGAGAGAGAGAGAAATACTACTTCTCAAAATGAGAGCCCTTAAG 219
169 laAsnLysProGlnAlaSerLeuGlnArg.....Gln 179
220 TGCTTGTACTGAAGACACAAATGTGCGCGTCA 255
180 CysValAsnLeuGlnAsnGlnLysGlyThrLysPro 191
```

seq_name: A_Geneseq_36:W39158

seq_documentation_block:

ID W39158 standard; Protein; 229 AA.

AC W39158; (first entry)
 DT 27-APR-1998
 DE Clone PRBS3FH2910 #3.1 CFH related protein fragment.
 KW Complement factor H; tumour associated antigen; renal cancer;
 OS urogenital cancer; medicament; modulator.
 PN Synthetic.
 PD WO9738136-A1.
 PF 16-OCT-1997.
 PR 09-APR-1997; US-812481.
 PR 06-MAR-1997; US-015083.
 PR 09-APR-1996; US-015083.
 PR 09-APR-1996; US-630048.
 PR 06-MAR-1997; US-038614.
 PA (BARD-) BARD DIAGNOSTIC SCI INC.
 PI Enfield DL, Hass GM, Kinders RJ;
 DR WPI: 97-512742/47.
 DR N-PSDB; V02794.
 PT Treating or screening for cancer, e.g. renal or urogenital cancer -
 PT by modulating or detecting tumour associated human complement factor
 PT H related antigen, or nucleic acid encoding it
 PS Example 6B; Fig 7B; 104pp; English.
 CC This partial protein is found in clone PRBS3FH2910 #3.1 and represents a
 CC human tumour-associated complement factor H (CFH). The detection of this
 CC protein and a CFH antigen can be used in screening or for the treatment
 CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.
 CC Agents that may modulate this antigen could be used in the manufacture of
 CC a medicament for the treatment of a tumour cell.
 SO Sequence 229 AA;

alignment_scores:
 Quality: 57.50 Length: 90
 Ratio: 1.223 Gaps: 5
 Percent Similarity: 52.222 Percent Identity: 27.778

alignment_block:
 US-09-030-606-224 x W39158 ..

Align seg 1/1 to: W39158 from: 1 to: 229

```

82 AGGAGTGTGTACCTACTCCCATCAAAAAAATGATCATGCT 131
   |||:::  ::::|  ::::|  ::::|  ::::|  ::::|
99 AAGTATGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 115
   |||:::  ::::|  ::::|  ::::|  ::::|  ::::|
132 T.....AAGATATGAAGGCAATATTTATCATATGCTC 166
   |||:::  ::::|  ::::|  ::::|  ::::|  ::::|
115 LserATGCTMetSerIysTyrProSerGlyGluArgValArgTyrGlnC 132
   |||:::  ::::|  ::::|  ::::|  ::::|  ::::|
167 TAAAGAG.....AAGAGAGAGAAATCTACTTCT 198
   |||:::  ::::|  ::::|  ::::|  ::::|  ::::|
132 ysATGSerProTyrGlnMetPheGlyAspGluValMetCysLeuAsn 148
   |||:::  ::::|  ::::|  ::::|  ::::|  ::::|
199 CRAATGGAAGC...CCTTAAAGTCTTGATGACTGAAGACAAATG 245
   |||:::  ::::|  ::::|  ::::|  ::::|  ::::|
149 GlysAnrPrpGluProGlnCysLysAspSerThrGly...LysCys 164
   |||:::  ::::|  ::::|  ::::|  ::::|  ::::|
246 TGGCCGCGCATCCCTTARAGTTGCATGACTGACACGCTAAGCTT 295
   |||:::  ::::|  ::::|  ::::|  ::::|  ::::|
164 sGlyProPro.PioProIleAsnGlnLysPleIleThrSerPheProLeu 180
   |||:::  ::::|  ::::|  ::::|  ::::|  ::::|
296 GCAGTTTARACTCMGCA 313
   |||:::  ::::|  ::::|  ::::|  ::::|  ::::|
181 SerValTyrIleAlaProAla 186
   |||:::  ::::|  ::::|  ::::|  ::::|  ::::|

```

seq_name: A_Geneseq_36:W28143

seq_documentation_block:

ID W28143 standard; Protein: 140 AA.
 AC W28143;
 DT 27-APR-1998 (first entry)
 DE Amino acid sequence of ATP dependent nuclease subunit B.
 KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;

KW Staphylococcal gene; regulatory element; bacterial gene expression;
 KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
 KW toxic shock syndrome.
 OS Staphylococcus aureus.
 PN WO9730070-A1.
 PD 21-AUG-1997.
 PF 19-FEB-1997; U02318.
 PR 20-FEB-1996; US-011888.
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PI Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,
 PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
 DR WPI: 97-424969/39.
 DR N-PSDB; T84068.
 PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
 PT to isolate antimicrobial compounds, and in vaccines against S.
 PT aureus infection
 PS Claim 6; Page 493; 989pp; English.
 CC The present sequence represents a staphylococcus aureus protein, that,
 CC based on homology with a Bacillus subtilis protein, is believed to be a
 CC ATP dependent nuclease subunit B. The DNA sequence was isolated from a
 CC library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA
 CC sequence can be used in the construction of ribozymes and antisense
 CC sequences to control the expression of staphylococcal genes. The DNA
 CC sequence is also useful as a source of regulatory elements for the
 CC control of bacterial gene expression. The present protein may be used
 CC to produce vaccines to enable a host to produce specific antibodies
 CC with antibacterial action. These vaccines and antibodies would protect
 CC a host against invasion by S. aureus, and conditions relating to
 CC Staphylococcal infection, e.g. aureus, and conditions relating to
 CC skin syndrome, and toxic shock syndrome.
 SO Sequence 140 AA;

alignment_scores:
 Quality: 57.00 Length: 76
 Ratio: 1.500 Gaps: 2
 Percent Similarity: 50.000 Percent Identity: 26.316

alignment_block:
 US-09-030-606-224 x W28143 ..

Align seg 1/1 to: W28143 from: 1 to: 140

```

86 AGTGTGACCTCTACTCCCATCAAAAAAATGATCATGCTTAA 135
   |||:::  |||  |||  |||  |||  |||  |||
37 SerIleuProLeuTyrAsnIleProTyrAsnIleSphr..... 50
   |||:::  |||  |||  |||  |||  |||  |||
136 GCATARAAGGCAATATTTATCATATGTTCTTAAAGAGAGAGAA 185
   |||:::  |||  |||  |||  |||  |||  |||
51 ...LysArgSerMetThrIshisPro..... 58
   |||:::  |||  |||  |||  |||  |||  |||
186 AATACTACTTCTCRAAATGGAAGCCCTTAAGGCTCTGATGGAAG 235
   |||:::  |||  |||  |||  |||  |||  |||
59 .....ValMetGluMetIleArgSerLeuIleGluValIle 70
   |||:::  |||  |||  |||  |||  |||  |||
236 GACACAATGTGGCGCATCCCTCTTARAGTTGCATGACTGACAC 285
   |||:::  |||  |||  |||  |||  |||  |||
71 GlnSerAsnTrpGlnValAsnProMetLeuArgLeuLeuLysThrAsp 87
   |||:::  |||  |||  |||  |||  |||  |||
286 GGTACTGTTCAGTTTARACTCMGCA 313
   |||:::  |||  |||  |||  |||  |||  |||
87 IleuThrAlaSerTyrIleuLysSerAla 96
   |||:::  |||  |||  |||  |||  |||  |||

```

seq_name: A_Geneseq_36:W39156

seq_documentation_block:

ID W39156 standard; Protein: 290 AA.
 AC W39156;
 DT 27-APR-1998 (first entry)
 DE Human partial complement factor H protein fragment 2.
 KW Complement factor H; tumour associated antigen; renal cancer;
 KW urogenital cancer; medicament; modulator.
 OS Homo sapiens.

PN W09738136-A1.
 PD 16-OCT-1997.
 PE 09-APR-1997; U05710.
 PR 06-MAR-1997; US-812481.
 PR 09-APR-1996; US-015083.
 PR 09-APR-1996; US-630048.
 PR 06-MAR-1997; US-038614.
 PA (BARD-) BARD DIAGNOSTIC SCI INC.
 PI Entfield DL, Hass GM, Kinders RJ;
 DR WPI: 97-512742/47.
 DR N-PSDB: V02792.
 PT Treating or screening for cancer, e.g. renal or urogenital cancer -
 PT H related antigen, or nucleic acid encoding it
 PS Example 6B: Fig 7B: 104pp: English.
 CC This partial protein sequence represents a region of the human
 CC tumour-associated complement factor H (CFH). This sequence is used in the
 CC identification of complement factor H related proteins and antigens from
 CC clone PRS33FH2910 (see W39157-W39159). The detection of this protein and
 CC a CFH antigen can be used in screening or for the treatment of renal or
 CC urogenital cancer, e.g. bladder, cervical or prostate cancer. Agents that
 CC may modulate this antigen could be used in the manufacture of a
 CC medicament for the treatment of a tumour cell.
 SO Sequence 290 AA;

alignment_scores:
 Quality: 57.50 Length: 90
 Ratio: 1.223 Gaps: 5
 Percent Similarity: 52.222 Percent Identity: 27.778

alignment_block:
 US-09-030-606-224 x W39156 ..

Align seg 1/1 to: W39156 from: 1 to: 290

```

82 AGGAGTGTGTACTCCCTTACTCCCATCAAAAAAATGATACATGCT 131
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
141 AYGAPThSerCysValAsnProthrvAlGlnAsnAlaTyrLleVa 157
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
132 T.....AAAGATARAAGGCATATTATTCATATGATGTC 166
   |      |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
157 LserArgIleMetSerLysTyrProSerGlyGluArgValArgTyrGlnC 174
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
167 TAAAGAG.....AAGGAGAGAAATACACTTCTTCT 198
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
174 yArgSerProTyrGluMetPheGlyAspGluGluValMetCysLeuAsn 190
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
199 CRAAATGGAAGC...CCTTAAAGTGCTTGTGATGAGACACAAATG 245
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
191 GYAAsnTrpThrGluProGlnCysLysAspSerThrGly..LysC 206
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
246 TGCCGCTCATCCCTTARAGTTGCATGACTTGACACGGTAACTGTT 295
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
206 sGlyProPro. PropProIleAspAsnGlyAspIleThrSerPheProLeu 222
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
296 GCAGTTTARACTCKGA 313
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
223 SerValTyrAlaProAla 228
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

```

seq_name: A_Geneseq_36:R56285

seq_documentation_block:

ID R56285 standard; Protein; 461 AA.
 AC R56285;
 DT 03-MAR-1995 (first entry)
 DE Sequence of human alpha-L-fucosidase including signal peptide.
 KW Anti-adhesion enzyme; alpha-L-fucosidase.
 OS Homo sapiens.
 FH Key
 FT peptide
 FT 1..22
 FT /label- signal
 FT 23..461
 FT protein

FT W09414472-A.
 PN 07-JUL-1994.
 PD 21-DEC-1993; U12464.
 PE 22-DEC-1992; US-994650.
 PR (REGC) UNIV CALIFORNIA.
 PA Carson DA, Wicks I;
 PI WPI: 94-234351/28.
 DR N-PSDB: 065537.
 PT Inhibiting adhesion of cells to selectin-contg. receptors - by
 PT treatment with enzyme degrading carbohydrate residues in selectin
 PT adhesion ligands, partic. for treating inflammation
 PS Disclosure; Page 27-30; 42pp: English.
 CC The AA sequence in R56285 is derived from sequence information obtd.
 CC through isolation of several human alpha-L-fucosidase cDNA clones
 CC reported by Occhiodoro et al. at Blochem. Biophys. Res. Comm.
 CC 164:439-445 (1989). The full-length nucleotide sequence for alpha-L-
 CC fucosidase is available from the GENBANK molecular sequence database
 CC under Accession No. M28099. The DNA sequence in Q65537 is described
 CC as genomic.
 SO Sequence 461 AA;

alignment_scores:
 Quality: 57.50 Length: 64
 Ratio: 1.597 Gaps: 2
 Percent Similarity: 56.250 Percent Identity: 28.125

alignment_block:
 US-09-030-606-224/rev x R56285 ..

Align seg 1/1 to: R56285 from: 1 to: 461

```

238 GTCTTCAAGTATCAAGACCTTTAAGGCTTCATTTTGAGAAAGTACT 189
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
189 LeuLeuAspLysLysAsnGlyPheLysThrGlnHisPheValSerAlaL 205
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
188 ATTTCCTCTCTCTCTTTTGAACATATGATAAATATGCGCTTTTA 139
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
205 S...ThreMetProGluLeuTyrAspLeuValAsnSerTyrLysProAspL 221
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
138 TCCTTAAACCATGATTCATTTTCTTTTGTGAGGAGGAGTAAAGGCTAC 89
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
221 eulle.....TrpSerAspGlyGluTyrIlePheLysPro 231
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
88 ACTCCCTACTACAAATGATCAAAACTTTTCTTTTGTGTTCT 47
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
232 AspThrTyrTrpAsnSerThrAsnPheLeuSerTyrPheLys 245
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

```

• • • •


```

205 .....CAT 202
30 sasprtmctvalphevalargglyprocluhstsserainleglnhisp 47
201 tv...GAGAAAGTAGATATTTCTCTCTCTCTTTAGACATATGATA 155
47 hevalgluluyvalvalpnehlstleuhsigluSerpheproargprolys 63
154 AATATGCGCCCTTATATCCTTTAACCATATATCATTTTTCATGCG 105
64 Argvalys..... 66
104 GGAGTAGAGGGTACACACTCCCTACTACATATGTCACAACTTTTCTT 55
67 .....LysaspproptotyrlsyalglugluSerglytyr 79
54 TGTGTCTCTATTTGGTTGTAAGTATTTTCTTACACAGAG 11
79 laglypheleleuprolleugluvaltyrPhehlysalnlysglu 93
seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-588-983-16

```

```

seq_documentation_block:
: Sequence 16, Application US/08588983
: Patent No. 5854067
: GENERAL INFORMATION:
: APPLICANT: Christopher B. Newgard, et al.
: TITLE OF INVENTION: Methods and Compositions
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Arnold, White & Durkee
: STREET: P. O. Box 4433
: CITY: Houston
: STATE: TX
: COUNTRY: US
: ZIP: 77210
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/588,983
: FILING DATE: Concurrently herewith
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Fussey, Shelley P.M.
: REGISTRATION NUMBER: 39,458
: REFERENCE/DOCKET NUMBER: UTSD:424/FUS
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312) 418-3000
: TELEFAX: (512) 474-7577
: TELEX: n/a
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 917 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-588-983-16

```

```

alignment_scores:
Quality: 56.50 Length: 67
Ratio: 1.527 Gaps: 5
Percent Similarity: 55.224 Percent Identity: 29.851

```

```

alignment_block:
US-09-030-606-224/rev x US-08-588-983-16 ..
Align seg 1/1 to: US-08-588-983-16 from: 1 to: 917

```

```

214 AAGGCTTCATTTGAGAAAGTAGATATTTCTCTCT..... 176
551 ArgglyvalgluMeHisAnlysletyrSerleprroglnguvalme 567
175 .....TCTCTTTAGAACATATGATAAATATGCCCC 142
567 thlsiglythrglugluuSerpheasphisilevalgl..... 580
141 YATATCTTAACATATATCATTTTTCG.....ATGGGAGATA 98
561 .....Cysile,Alasphleugluurymetglymety 592
97 GGGGTACACACTCCCTACTACATATGTCACAACTTTTCTTTTGT 49
592 scllyalserleupro.....leuglyphethPheSerpheprolys 606

```

```

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-07-688-352C-40
seq_documentation_block:
: Sequence 40, Application US/07688352C
: Patent No. 5527896
: GENERAL INFORMATION:
: APPLICANT: Wiegler, Michael H.
: APPLICANT: Colicelli, John J.
: TITLE OF INVENTION: Cloning by Complementation and Related
: NUMBER OF SEQUENCES: 57
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
: STREET: Two First National Plaza, 20 South Clark
: STREET: Street
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60603
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/688,352C
: FILING DATE: 19910419
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/511,715
: FILING DATE: 20-APR-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: Borun, Michael F.
: REGISTRATION NUMBER: 25447
: REFERENCE/DOCKET NUMBER: 27805/30197
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312) 346-5750
: TELEFAX: (312) 984-9740
: TELEX: 25-3856
: INFORMATION FOR SEQ ID NO: 40:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 384 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-07-688-352C-40

```

```

alignment_scores:
Quality: 55.00 Length: 85
Ratio: 1.222 Gaps: 5
Percent Similarity: 52.941 Percent Identity: 30.588

```

```

alignment_block:
US-09-030-606-224/rev x US-07-688-352C-40 ..

```

Align seg 1/1 to: US-07-688-352C-40 from: 1 to: 384

```

290 TTACCGTGTCAGATGATGACACTTAAAGAGATGAGCGGCCCATTT 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
305 VAIProValProLeuHisSerSerProHisSerProAlaProProThr 321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
240 GTGTCCTTCATCAAGACACCTTTAAGGCGCTCCATTT 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
321 rSerSerSerLysSerLysThrLeuValLeuLeuSerLeuLeuArgLysL 338
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
200 ....YGAAGAAGTATGATTTCTCTCTCTCTCTTTAGACATATGAT 156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
338 yslYstHlySerGlyValPhePheCysPheLeuPhe..... 350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155 AAATATATGCCCTTATATCTTTA...AGCATGATCATTTTTTTGTG 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
351 ....PheProPheProProAlaProThrHisGlyAlaPhePheThrPar 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
108 ATGGGAGATAGGGGTACACACTCCCTACTACATGTCACAACTTTTTC 59
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
365 gtrGlyLeu...GlyAsnGlyGly...LeuArgSerArgLysGlyPheT 380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
58 TTTT 54
    |||
380 yrphe 381
    |||

```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-446-875-10

```

seq_documentation_block:
; Sequence 10, Application US/08446875
; Patent No. 5858751
; GENERAL INFORMATION:
; APPLICANT: Paulson, James C.
; APPLICANT: Wen, Xiaohong
; APPLICANT: Livingston, Brian Duane
; APPLICANT: Gillespie, William
; APPLICANT: Kelm, Sorge
; APPLICANT: Burlingame, Alma L.
; APPLICANT: Medzinski, Katalin
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Identification and Synthesis of Sialyltransferases
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Poms, Smith, Lande & Rose
; STREET: 2029 Century Park East, Suite 3800
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,875
; FILING DATE: July 12, 1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/102,385
; FILING DATE: August 4, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Oldenkamp, David J.
; REGISTRATION NUMBER: 29,421
; REFERENCE/DOCKET NUMBER: 111-197
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (310) 788-5000
; TELEFAX: (310) 277-1297
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 375
; TYPE: amino acid

```

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-446-875-10

alignment_scores:
Quality: 54.50 Length: 91
Ratio: 1.329 Gaps: 5
Percent Similarity: 45.05 Percent Identity: 27.473

alignment_block:
US-09-030-606-224/rev x US-08-446-875-10 ..

Align seg 1/1 to: US-08-446-875-10 from: 1 to: 375

```

315 AATCGKAGTTTAAACTGCAACAGTTACCGTGTCAGATGACACTY 266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65 AsnLeuSpSerLysLeuProAlaGluLeuAlaThrLysTyrAlaAsnPh 81
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
265 TAAAGAGAGA.....TGACGCGCA 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
81 eSerGluGlyAlaCysLysProGlyTyrAlaSerAlaLeuMethTrAlaI 98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
245 CATTTGTCTCTCAGTATCAAGACACCTTTAAGGCGCTCCATTTYGAGA 196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98 LepProArgPheSerLysProAlaProMet..... 108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
195 AAGTAGATTTCTCTCCCTCTCTTTAGACATGATGATAAATATGTC 146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
109 .....PheLeu.AspAspSerPheArgLysTyrAlaArgLysLeuArg 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 CTTTATCTCTTAACCATGATTCATTTTGTGATGGGAGTAAAG 96
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
122 LnpheValPro.....ProPhe.....GlyIleLys 130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 GGTACACACTCCCTACTACAA 75
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
131 GlyLnsAspAsnLeuIleLys 137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS9_COMB.pcp:US-08-03916-17

```

seq_documentation_block:
; Sequence 17, Application PC/TUS9603916
; GENERAL INFORMATION:
; APPLICANT: Wild, Martha A.
; APPLICANT: Cochran, Mark D.
; TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/03916
; FILING DATE: 23-MAR-1995
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/126,597
; FILING DATE: 24-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 39116-A
; TELECOMMUNICATION INFORMATION:

```

TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 292 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US96-03916-17

alignment_scores:
Quality: 54.00 Length: 66
Ratio: 1.588 Gaps: 4
Percent Similarity: 51.515 Percent Identity: 27.273

alignment_block:
US-09-030-606-224/rev x PCT-US96-03916-17 ..

Align seg 1/1 to: PCT-US96-03916-17 from: 1 to: 292

190 GTATTTCCTCTCTCTTTAGACATATGATAAATATGCCCTT 141
:::||||| |::: |::: |
207 LeuphSerThrArgSerLeuSerLysGlnArgLeuArgPheCysAsp.L 223
140 TA.....TCCTTAACCATGTA 124
223 euArgLeuSerLysSerArgLeuPheSerThrArgSerLeuSerLys 239
123 TCCATTTTTCGATGGAGGAGTACACACTCCCTACT...AC 77
||||| |::: |::: |::: |
240 ProArgPheLeuThrLeuGlyProArgGlyPheArgLeuGlyThrArgTh 256
76 AATGCAACAACCTTTCTTTTGT.....TGTTCCTAT 43
|::: |::: |::: |::: |
256 rLeuSerLysAspHisArgPheCysThrLeuGlyLeuCysSerPhe 271
seq_name: /cgn2_6/ptodata/2/1aa/5b_COMB.pep:US-08-726-525-4

seq_documentation_block:
Sequence 4, Application US/08726525
Patent No. 5789181
GENERAL INFORMATION:
APPLICANT: Lln, Lln-Ling
TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726,525
FILING DATE: 07-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,942
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15258
TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 320 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-726-525-4

alignment_scores:
Quality: 53.50 Length: 69
Ratio: 1.337 Gaps: 3
Percent Similarity: 57.971 Percent Identity: 31.884

alignment_block:
US-09-030-606-224 x US-08-726-525-4 ..

Align seg 1/1 to: US-08-726-525-4 from: 1 to: 320

47 GAACAACAAGAAAAA.....GTTTGACATTGTAGGAGAGTGT 90
|::: |::: |::: |::: |
166 GlutylGlnLysLysGlnGlyLleValCysLysGlnGlnValGlyGlyVa 182
91 GTACCCCTACTCCCAACAAAAAATGATACATGCTTAAGATA 140
|::: |::: |::: |::: |
182 lLeuGlnLeuPheProIleAsnGlySerSerValValGlnArgGlnAspV 199
141 RAAGGCATATTTTATCATATGTTCTAAAGAGAGAGAGAGAAATAC 190
||| |::: |::: |
199 alproAla.....HisLeuVal.....LysAspIleArgAsnTyr 210
191 TACTTTCCTAATGAGAGCCCTTAAGGTCCTTGATACTGAAGACAC 240
::: |::: |::: |::: |
211 PheGlnValSerProGlnTyrPheSerMetLeuLeuValGlyLysAspG1 227
241 AATGTG 247
||| |
227 yAsnVal 229

seq_name: /cgn2_6/ptodata/2/1aa/5b_COMB.pep:US-08-487-942-4

seq_documentation_block:
Sequence 4, Application US/08487942
Patent No. 5817476
GENERAL INFORMATION:
APPLICANT: Lln, Lln-Ling
TITLE OF INVENTION: NOVEL INTERLEUKIN-1 RECEPTOR
TITLE OF INVENTION: INTRACELLULAR LIGAND PROTEINS AND INHIBITORS OF LIGAND
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: LEGAL AFFAIRS, GENETICS INSTITUTE, INC.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,942
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15258

;; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/062,443

```

;
; APPLICANT: Croce, Carlo
; APPLICANT: Canaanl, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods

```

```

? TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
? TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1
? NUMBER OF SEQUENCES: 86
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &
? ADDRESSEE: Norris
? STREET: One Liberty Place, 46th floor
? CITY: Philadelphia
? STATE: Pennsylvania
? COUNTRY: USA
? ZIP: 19103
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US94/04496
? FILING DATE:
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: Deluca Esq., Mark
? REGISTRATION NUMBER: 33,229
? REFERENCE/DOCKET NUMBER: TJU-1242
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (215) 568-3100
? TELEFAX: (215) 568-3439
? INFORMATION FOR SEQ ID NO: 31:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 559 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? PCT-US94-04496-31

```

alignment_scores:

```

Quality: 54.00 Length: 114
Ratio: 1.174 Gaps: 5
Percent Similarity: 40.351 Percent Identity: 24.561

```

alignment_block:

```
US-09-030-606-224/rev x PCT-US94-04496-31 ..
```

```
Align seg 1/1 to: PCT-US94-04496-31 from: 1 to: 559
```

```

295 AACAGTACCGTGTCCAGTCATGCACTTTAAGGAGATGAC...CG 249
|||||  |||||||
3 AsnGlnCysThrValGlnVal.....ArgLeuGlnLeuG1 14
248 CCACATTGTGTCTTCAGATCAAGACCTTTAAGGGCTTC..... 206
|||||  |||||||
14 YHisArgIleGlnLeuArgLysProThrGlnGluGlyPheThrHisA 31
205 .....CATTT 200
31 sPTpMetValPheValArgGlyProGluGlnCysAspIleGlnHisPhe 47
|||||  |||||||
199 ...GAGAAAGTAGATTTTCTCTCTCTCTTTTGAACATATGATAA 153
|||||  |||||||
48 ValGlnLysValValPheThrPheHisAspSerPheProLysProArgAr 64
132 ATATTGCCCTTTATCTTAAACATGATTCATTTTTTTTGATGGGG 103
|||||  |||||||
64 gValLys..... 66
102 AGTAAGGGGTACACACTCCCTACTACAAATGTCACAAACTTTTCTTTTG 53
|||||  |||||||
67 .....LysGlnProProTyrLysValGlnGlnSerGlyTyrAla 79
|||||  |||||||
52 TTGTTCTATTGGTGTACTACTATTCTTACACAGAG 11
|||||  |||||||

```

```

80 GlyPheIleMetProIleGlnValHisPheLysAsnLysGlu 93
seq_name: /cgn2_6/prodata/2/1aa/5B_COMB.pep:US-08-806-581A-2

```

```

seq_documentation_block:
? Sequence 2, Application US/08806581A
? Patent No. 5780264
? GENERAL INFORMATION:
? APPLICANT: Wessling-Resnick, Marianne
? APPLICANT: Gutierrez, Jesus
? TITLE OF INVENTION: IRON TRANSPORT PROTEIN
? NUMBER OF SEQUENCES: 2
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: HALE AND DORR LLP
? STREET: 60 State Street
? CITY: Boston
? STATE: MA
? COUNTRY: United States
? ZIP: 02109
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/806,581A
? FILING DATE: 10-JUN-1996
? CLASSIFICATION: 514
? ATTORNEY/AGENT INFORMATION:
? NAME: Baker, Hollie L.
? REGISTRATION NUMBER: 31,321
? REFERENCE/DOCKET NUMBER: 426.97.177
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (617) 526-5000
? TELEFAX: (617) 526-6000
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 337 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE:
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? US-08-806-581A-2

```

alignment_scores:

```

Quality: 53.00 Length: 60
Ratio: 1.432 Gaps: 3
Percent Similarity: 61.667 Percent Identity: 31.667

```

alignment_block:

```
US-09-030-606-224/rev x US-08-806-581A-2 ..
```

```
Align seg 1/1 to: US-08-806-581A-2 from: 1 to: 337
```

```

283 GTCCAGATGATGACACTTTAAGGAGATGAGCGGCACATTTGTGTCCT 234
|||||  |||||||
146 LeuGlnIleGlnGlnThrCysLysArgMetAspSerGluIleCysAsnVa 162
233 TCAGTATCAAGACACCTTTAAGGGCTTCATTTTGAGAAAGTACTATTTT 184
|||||  |||||||
162 IlyIleLysAsnIlePhe.....Ile..TyrAsnGlnSerThrValLys 176
183 CTCTTCCTCTCTTTTGAACATATGATTAATATGCCCCCTTAT... 138
|||||  |||||||
176 sSerArgPhePhePheAsnIleSerSerLysLysLeuSerThrValTyrLeuP 193
|||||  |||||||
137 ..CCTTAACCATGATTCATTTT 111
|||||  |||||||
193 heGlyCysThrMetLysProPheAspPhe 202

```

```
seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-602-010A-30
seq_documentation_block:
; Sequence 30, Application US/08602010A
; Patent No. 5753235
; GENERAL INFORMATION:
; APPLICANT: Haanes, Elizabeth J.
; APPLICANT: Frank, Rexann S.
; TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,010A
; FILING DATE: February 15, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-46
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/863-9700
; TELEFAX: 303/863-0223
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-602-010A-30

alignment_scores:
Quality: 51.00 Length: 39
Ratio: 1.889 Gaps: 1
Percent Similarity: 69.231 Percent Identity: 28.205

alignment_block:
US-09-030-606-224/rev x US-08-602-010A-30 ..
Align seg 1/1 to: US-08-602-010A-30 from: 1 to: 53

253 GACGGCCACATTTGTGCTTCAGTATCAAGCACCTTAAGGCTTCCA 204
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10 Aspelnhisleuservalasnasntyrserglythrileglnphelel 26
203 TTYGAGAAAGTAGTATTTCCTCTCTCTCTTTAGACATATGATAA 154
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26 sphaasnasnsercystyrthrvaltyrglnthrilleglutyrphe.... 41
153 AATATTGCCCTTYATC 137
|||||:|||||:|||||:|||||:|||||:|||||:
42 ..Sercysproargile 46

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-680-726A-30
seq_documentation_block:
; Sequence 30, Application US/08680726A
; Patent No. 5804197
; GENERAL INFORMATION:
; APPLICANT: Haanes, Elizabeth J.
; APPLICANT: Frank, Rexann S.
; TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
```

```

; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,726A
; FILING DATE: 12-JUL-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-46-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-680-726A-30

alignment_scores:
Quality: 51.00 Length: 39
Ratio: 1.889 Gaps: 1
Percent Similarity: 69.231 Percent Identity: 28.205

alignment_block:
US-09-030-606-224/rev x US-08-680-726A-30 ..
Align seg 1/1 to: US-08-680-726A-30 from: 1 to: 53

253 GACGGCCACATTTGTGCTTCAGTATCAAGCACCTTAAGGCTTCCA 204
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10 Aspelnhisleuservalasnasntyrserglythrileglnphelel 26
203 TTYGAGAAAGTAGTATTTCCTCTCTCTCTTTAGACATATGATAA 154
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26 sphaasnasnsercystyrthrvaltyrglnthrilleglutyrphe.... 41
153 AATATTGCCCTTYATC 137
|||||:|||||:|||||:|||||:|||||:|||||:
42 ..Sercysproargile 46

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-465-980-3
seq_documentation_block:
; Sequence 3, Application US/08465980
; Patent No. 5756309
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel R.
; APPLICANT: Li, Yi
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPRAJ70
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
; STREET: STUART & OLSTEIN
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 12:41:49 ; Search time 2928.69 Seconds
(without alignments)
347.493 Million cell updates/second

Title: US-09-030-606-224
Perfect score: 320
Sequence: 1 CCCCGAAGCGCTTCTGTAA.....TTTARACTCMGCATTGTGAC 320

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database

1:	genEmb1.*
2:	gb_ba1.*
3:	gb_ba2.*
4:	gb_om.*
5:	gb_ov.*
6:	gb_pat.*
7:	gb_ph.*
8:	gb_pl1.*
9:	gb_pl2.*
10:	gb_pr1.*
11:	gb_pr2.*
12:	gb_ro.*
13:	gb_st.*
14:	gb_st.*
15:	gb_sy.*
16:	gb_un.*
17:	gb_v.*
18:	em_fun.*
19:	em_hb.*
20:	em_hun1.*
21:	em_hun2.*
22:	em_in.*
23:	em_om.*
24:	em_ov.*
25:	em_ov.*
26:	em_pat.*
27:	em_ph.*
28:	em_pl.*
29:	em_ro.*
30:	em_st.*
31:	em_sy.*
32:	em_un.*
33:	em_v.*
34:	gb_hg1.*
35:	gb_hg2.*
36:	gb_in1.*
37:	gb_in2.*
38:	em_ba1.*
39:	em_ba2.*
40:	em_hun3.*
41:	em_hun4.*
42:	gb_pr.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	45	14.1	10094	2	067573	067573 Methanococcus

2	43.8	13.5	1.99917	34	AC006755	Caenorhab
C 3	43.2	13.5	1.4553	37	AE001391	AE001391 Plasmodium
C 4	40.8	12.7	1.94410	35	AC005140	Plasmodium
C 5	40.8	12.7	2.19200	35	AC006279	Plasmodium
C 6	40.4	12.6	4.013	1	SDNAWPR	Plasmodium
7	40.4	12.6	2.9777	36	CEFA6C3	Plasmodium
8	40	12.5	1.92151	35	AC007282	Plasmodium
C 9	39.8	12.4	7.8338	42	AF129078	Plasmodium
C 10	39.6	12.4	9.1029	34	HS72817	Plasmodium
C 11	39.6	12.4	3.12766	34	PFMA1AP3	Plasmodium
C 12	38.8	12.1	8.9603	8	AC007659	Plasmodium
C 13	38.8	12.1	8.9534	8	ATAC002398	Plasmodium
C 14	38.8	12.1	6.9037	12	AC003949	Plasmodium
C 15	38.4	12.0	1.38888	10	HS800F24	Plasmodium
C 16	38.4	12.0	2.61670	34	PFMA1AP2	Plasmodium
17	38.4	12.0	1.21336	42	AC006458	Plasmodium
C 18	38.2	11.9	7.218	5	166494	Sequence 14
C 19	38.2	11.9	8.7104	8	ATF5C23	Arabidops
20	38.2	11.9	3.24429	36	CEP67H2	Arabidops
21	38.2	11.9	2.3952	36	CEP29D10	Arabidops
22	38	11.9	1.0214	2	MCU62057	Arabidops
23	38	11.9	1.26041	11	AC004904	Arabidops
24	38	11.9	1.90434	11	AC005771	Arabidops
25	37.8	11.8	1.04871	11	AC004584	Arabidops
26	37.8	11.8	1.16019	11	AC004629	Arabidops
27	37.8	11.8	5.51	36	DDCOXV	Arabidops
C 28	37.8	11.8	8.1023	37	AC001659	Arabidops
C 29	37.6	11.8	8.4440	7	AB010075	Arabidops
C 30	37.6	11.8	1.18086	7	ATF6H11	Arabidops
C 31	37.6	11.8	1.80000	35	AC004578	Arabidops
C 32	37.6	11.8	1.90165	35	AC006235	Arabidops
C 33	37.6	11.8	1.80117	35	AC007053	Arabidops
C 34	37.6	11.8	1.65283	37	AC005711	Arabidops
C 35	37.6	11.8	2.03199	37	AC005721	Arabidops
36	37.4	11.7	4.3441	1	MYL1PBC	Arabidops
37	37.4	11.7	1.17381	11	AC006392	Arabidops
C 38	37.2	11.6	1.17056	3	MDNDCOMGN	Arabidops
C 39	37.2	11.6	1.32785	36	CE145P10	Arabidops
40	37.2	11.6	3.9781	36	CE14ZK418	Arabidops
41	37	11.6	1.1246	2	U67600	Arabidops
42	37	11.6	7.1934	9	HS159A15	Arabidops
43	36.8	11.5	1.10000	34	CEX11B2-0	Arabidops
C 44	36.8	11.5	2.20000	36	AC007126	Arabidops
C 45	36.8	11.5	1.64399	36	PFMA13P6	Plasmodium
298511						Plasmodium

ALIGNMENTS

RESULT	1
LOCUS	U67573
DEFINITION	U67573 10094 bp DNA BCT 20-MAY-1998
ACCESSION	Methanococcus jannaschii section 115 of 150 of the complete genome
NID	U67573 L77117
VERSION	92826395
KEYWORDS	U67573.1 GI:2826395
SOURCE	.
ORGANISM	Methanococcus jannaschii.
REFERENCE	Methanococcus jannaschii.
AUTHORS	Archaea, Euryarchaeota; Methanococcales; Methanococcaceae;
	Methanococcus.
	1 (bases 1 to 10094)
	Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D.,
	Sutton,G.G., Blake,J.A., Fitzgerald,L.M., Clayton,R.A.,
	Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J., Adams,M.D.,
	Reich,C.I., Overbeek,R., Kirkness,E.F., Weinstock,K.G.,
	Merriick,J.M., Glodek,A., Scott,J.D., Geoghegan,N.S., Weidman,J.F.,
	Fuhrman,J.L., Nguyen,D.T., Uitterback,T., Kelley,J.M.,
	Peterson,J.D., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A.,
	Roberts,K.M., Kaine,B.B., Borodovsky,M., Klenk,H.P., Fraser,C.M.,
	Smith,H.O., Woese,C.R. and Venter,J.C.
	Complete genome sequence of the methanogenic archaeon,
	Methanococcus jannaschii


```

/transl_table=1
/product="GTP-binding protein, member of GTP1/OBG-family"
/protein_id="AAB99340.1"
/db_xref="PID:g1591973"
/db_xref="GI:1591973"
/translation="MAATIGLVKRPVNGKSTWENALTEKPAEIGNYPTTIOPNKGLAY
ITSPCKEGLVKCNPNRNSCIDGIRHPIPEVIDAGLVGAGHGRGMRKFLDLDLQ
ADAFILVVDASGKIDAGNPNEDPVEDKFLNEIDMTYSILITNMRLARRAQ
EKNIKVALKQOLSGINDEDIDKMAIRDMSPIKMTEDDLNATKRLKRSKPMITA
KNEADHPDARADYIKDFLKKYGGTGVQECINKATFDLDMTVIPEDEKNSFSDQGNV
LPDAFLVKGSTANDLAVKVTLEGEKFIYAIDAKKIRIVGADYELKHNDIIRIVSNA
K"
/gene="7436"
/gene="MJ1333"
/182..7436
/7182..7436
/gene="MJ1333"
/note="hypothetical protein; identified by Genemark;
putative"
/transl_table=1
/codon_start=1
/product="M. jannaschii predicted coding region MJ1333"
/protein_id="AAB99345.1"
/db_xref="PID:g1591974"
/db_xref="GI:1591974"
/translation="MSPKIOEITNELDNIMNRKYLELVATVEYLLNLIDPSKREK
FKELVDATVEVEDYELIKAIKIDLGMOGARVYLITLEGQ"
/gene="7908"
/complement(7495..7908)
/gene="MJ1333.1"
/complement(7495..7908)
/note="Brute Force ORF; identified by Genemark; putative"

Query Match      14.18; Score 45; DB 2; Length 10094;
Best Local Similarity 48.28; Pred. No. 0.38; Mismatches 127; Indels 0; Gaps 0;
Matches 120; Conservative 2;

QY      5 TGAAGGCTTCTGTGTAAGTAAGTACAGTACACCAATGGAACACAAAGAAAA 64
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3438 TGAAGACACTAAAGATGAGAAATTTTAAACCAAAATTCACAAATATATAAAA 3497
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      65 GTTGTGACATGTAGTAGGAGTGTGACCCCTTACTCCCATCAAAAAAATGAT 124
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3498 ATGTGTGAAGTTTAAAGACAGATAGATAGATACCTCCAAAAAACAATATGTT 3557
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      125 ACATGCTTAAGATARAAGGCATATTTATCATATGTTCTAAAGAGAAGAGAGA 184
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3558 TAATTGGCAGCATTGACAACTATTAATTTTATTAATAAATCTAAATAAAGCAT 3617
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      185 AATACACTACTTCTCRAAATGAAGCCCTTAAGGTGCTTGATCTGAAGACACAAAT 244
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3618 ATATGTTCCAACTTACACAGAAACCATATATATAGTTAGTAGAGATGATGAAC 3677
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      245 GTGGCCCTC 253
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3678 ATTAACCTTC 3686
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
LOCUS AC006755 199917 bp DNA HTG 23-FEB-1999
DEFINITION Caenorhabditis elegans clone Y40C5, WORKING DRAFT SEQUENCE, 1
ACCESSION AC006755
NID 94263183
VERSION AC006755.1 GI:4263183
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 199917)
AUTHORS Waterston,R.H.

```

```

TITLE The sequence of Caenorhabditis elegans clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 199917)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT * NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
source
1. 199917
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y40C5"
BASE COUNT 63887 a 33980 c 34366 g 65836 t 1848 others
ORIGIN
Query Match      13.78; Score 43.8; DB 34; Length 199917;
Best Local Similarity 51.78; Pred. No. 0.52; Mismatches 82; Indels 0; Gaps 0;
Matches 90; Conservative 2;

QY      30 ACATTAACCAACCAATAGACACCAAAAGAAAGTTGTGACATGTAGTAGGAGTG 89
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 127380 AAAATGAACTATTTGATGAGAAATGCGCAAACTTTGGCAGATTGGCCAGAGTTG 127439
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      90 TGAACCCCTTACTCCCATCAAAAAAATGATATACATGTTAAAGATARAAGGCAA 149
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 127440 TGTAACTTTTTCACGTTTAAACGAGTTATATACCTGGAATTAATGAACATTA 127499
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      150 TATTTATCATATGTTCTAAAGAGAAGAGAAATATCTTCTCRAA 203
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 127500 GTTATGCCAATATCTGTAATAAATTTACTTAACCAATTTATCTTGATTA 127553
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
LOCUS AE001391/c 14553 bp DNA INV 06-NOV-1998
DEFINITION Plasmodium falciparum chromosome 2, section 28 of 73 of the
ACCESSION AE001391 AE001362
NID 93845168
VERSION AE001391.1 GI:3845168
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 14553)
AUTHORS Gardner,M.J., Petzellin,H., Carucci,D.J., Cummings,L.M., Aravind,L.,
Koonin,E.V., Shalimov,S., Mason,T., Yu,K., Fujii,C., Pederson,J.,
Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Perlea,M.,
Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O.,
Smith,H.O., Fraser,C.M., Adams,M.D., Venter,J.C. and Hoffman,S.L.
Chromosome 2 sequence of the human malaria parasite Plasmodium
falciparum
JOURNAL Science 282 (5391), 1126-1132 (1998)
MEDLINE 99021743
ERRATUM: Erratum: [[published erratum appears in Science 1998 Dec
4;282(5395):1827]]
REMARK 2 (bases 1 to 14553)
REFERENCE Gardner,M.J.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-1998) The Institute for Genomic Research, 9712
Medical Center Drive, Rockville, MD 20814, USA
FEATURES
Location/Qualifiers

```

```

source
1. .14553
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="2"
/complement(202. .6234)
/gene="PFB0380c"
/complement(202. .6234)
/gene="PFB0380c"
/feature="PFB0380c"
/feature="identified by sequence similarity; putative"
/codon_start=1
/product="phosphatase (acid phosphatase family)"
/protein_id="AAC71865.1"
/db_xref="PDB:93845169"
/db_xref="GI:3845169"
/translation="MLIKQPEVKEKKEKKEKGAADKQDLPSLKKREKREKESOK
IDRYLINSKDSKNSKYSCYLNNECVKINSICKKCKKPEKNTKTYIRHAGART
PKKIKNIWPEKGGDLTFLGQOSIKVGEYLRKYTFENLKNKYNREKGLRLIN
KEGYIRKNCQVCKCTYKKNYNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
GYSYLDSTSEFLENRRKLKHPYLDPEFYEKYFLKINRSKHORVYKIKR
RRNNLIKIMIHOLINKKIKIKKNMNNYKCYIKESIRKRYKRMENICNN
DDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
DETFNKSENVADMKTYYKYNLKKRNKTYNNKKELPEPMEHLMMKRL
INKKRNKIRKKRKYDKIKLKNKYLCTKTENSEKCLAVGILGILGISEYTF
FLLFERSYNDKTDNNIDYTKRERKCKLNRKCFQNMILNRPDITSGOYCIDKN
TAVKNYIIGENLGENGCGKNGCDILGDLICGDLIGDNNISIPFRSNRIFCOS
KTFCEDELYFENKILRLOSIDDMYKINHEVMPGNDVLYNSYKCYDKNDGYSY
PSYKSNQYKSHVYIKKKNKYSVOCSESTIKREOREKKEKKEKKEKKEKKEK
NNMTNINWFDLILNERNFOFYNNIKKROKKNGLBENNVIRFOLIKYTLNE
ESKFEFLFKNKNVENIDTSTNTINIKYINNVARHDCFEKKQHSKEHMKK
IHLRDFIEYKENEIIDNCNNINNDKKEKKEKKEKKEKKEKKEKKEKKEKKEK
IYLLYNDYDYOFLFYLYKNTYILFVVKVARSMLKTLKTNHITKLNHI
HNSDYKILNNYKDELFYDDTKEWENKNTDILYDKKNTKIDLENDYDPII
TNDKEHYHNNNSIISYLKKNHNSVYLKKEKKEKNTDILKCLKCNKIKYHNNY
DKNKTYODKIRKNTYHPRHKKKNKIKKTSADATYHCVNMLNPNRAYELAS
QHPSSIDILKKEGONNTITGEIKYEONNFIKRPNTINSGLNCSLCHESNSN
TLAGNREANILADREDRKRNKQNRKQNGMTINNSKKNYNNNNNNNNNNNN
KKYDQKNDQTEOCQKNDQTEOCNDQTEOCNDQTEOCNDQTEOCNDQTEOCND
KKNYCYKLMCKEYKSNKYSWLCGSMGLDIVINEINVRLEYKENEKTKCFI
PRIILYTHOSSILSFOSGVRKIKRMKIPPEFSLSLEIHKKIKKIKNSKICN
SNNEKYSYKKNYKIMKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEK
VAKRYTEBNEKFEKCRKKNKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEK
NTYINDIPTVNNKRYESTLPRCLNKHIDFKMLFYLLCYKNNNIDQLDQICAN
NYTHIRKMQLEKGGHKGKNFYGVKFEFNKSVLCKLKKKIKKKNMGKNDKEE
DNNYHNDKNNYSDNIIFYNDHTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NANKRESLKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
TAKKQDENGSCINNRKRNKIKRNATRNILNNSNNKMYATYINSTNYTGKRYK
KIKKQVHTNNILHTYOKKRRKSTIISDNNNNNNNNNNNNNNNNNNNNNNNN
KOKYINDHNNINSDNNINNGLINERKVLHNDCKNNKNNQIIGYSIKRYKNSC
SDVTISLDRKIKRKKRKLQKKYENENIVCDLCLISYLKMLRIGNPEIL"
9488. .10522
/gene="PFB0385w"
/join(9488. .9512,10001. .10200,10376. .10522)
/gene="PFB0385w"
/feature="identified by sequence similarity; putative"
/codon_start=1
/product="acyl carrier protein"
/protein_id="AAC71866.1"
/db_xref="PDB:93845170"
/db_xref="GI:3845170"
/translation="MEVLSYVYVGSVLKRRSQVFNLRKNDYNIKKNPSS
LKSTPDIKIISKOLSVEDKIQMSNFKDGLADSLDLVELMALERKFNVTISDQ
DAKINTVDADIDYIERKNNQ"
11515. .13011
/gene="PFB0390w"
11515. .13011
/gene="PFB0390w"
/feature="identified by sequence similarity; putative"
/codon_start=1
/product="ribosome releasing factor (OR, RF)"
/protein_id="AAC71867.1"
/db_xref="PDB:93845171"
/db_xref="GI:3845171"
/translation="MIFQHLSLHNEKISMPFYTLILYISHVIRIFYDCLNLKNEK

```

```

RESULT 4
AC005140/c
LOCUS
DEFINITION
AC005140
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
On Apr 2, 1999 this sequence version replaced gi:4337162.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 13954: contig of 13954 bp in length
* 13955 14154: gap of unknown length
* 14155 32431: contig of 18277 bp in length
* 32432 32632: gap of unknown length
* 32632 105078: contig of 72447 bp in length
* 105078 105278: gap of unknown length
* 105278 117575: contig of 12297 bp in length
* 117575 117776: gap of unknown length
* 117776 120340: contig of 2565 bp in length
* 120340 120541: gap of unknown length
* 120541 132816: contig of 12276 bp in length
* 132816 133017: gap of unknown length
* 133017 135389: contig of 2373 bp in length

```

	a	c	g	t	others
BASE COUNT	88210	22418	19532	88036	1004
ORIGIN					

Query Match	12.7%;	Score 40.8;	DB 35;	Length 219200;
Best Local Similarity	47.2%;	Pred. No. 2.4;		
Matches 117;	Conservative 2;	Mismatches 129;	Indels 0;	Gaps 0;

Qy 20 AGAAATAGTACGTTACACCATAGACAACAAAAAGTTGTGACTTGTA 79
| | | | | | | | | | | | | | | |
Db 85299 ATAAAAAATTAAAAAATAAAAAATAAAAAATAAAAAATAAA 85240

80 GTAGGAGTGTACCCCTTACTCCCACTCAAAAAAAAAATGGATACATGTTAAGCAT 139

Out	DD	85253	AAA1	AAAAA1	AAAAAA1	AAAAAAA1	AAAAAAAA1	AAAAAAAGAT	85180
140	ABABCCCCC	AAATATTTT	ATCATATCT	CTCAAAA	CCGACG	AGAAA	ATATATCT	CTCTCTCT	100

D8 GGAGGATTAATTCTTGTATAAAATCAATACATATATAA **B51200**

200 RAAATGGAGCCCTTAAAGTCTTGTACTGAAGACACAATGTGGCCCTCCATCCT 259

Db 85119 TATTTATGCTAAATTT 85060

QY 260 CITTARA 267

Db 85059 TCCTTAA 85052

RESULTS

LOCUS	SADNAMDPB	4013 bp	DNA	BCT	01-AUG-1994
SADNAMDPB/c					

DEFINITION *S. aureus* plasmid encoded DNA, mup R gene.
ACCESSION X75439

NID 9438226
VERSION X75439.1 GT:438226

KEYWORDS Isoleucyl-tRNA synthetase.
SOURCE *Staphylococcus aureus*.

ORGANISM
Staphylococcus aureus
Bacteria: Firmicutes; Bacillus/Clostridium group; Bacillaceae;
Staphylococcus

REFERENCE
1 (bases 1 to 4013)
Hodgson, J. F., Curmick, S. P., Dyke, K. G., Morris, R., Sylvester, D. R., Scapnyiotoccus.

TITLE Molecular characterization of the gene encoding high-level and Gross, M.S.

mupirocin resistance in *Staphylococcus aureus* J2870
Antimicrob. Agents Chemother. 38 (5), 1205-1208 (1994)

MEDLINE	94346838
REFERENCE	2 (bases 1 to 4013)

AUTHORS	Direct Submission
TITLE	
Hodgson, J.E.	

JOURNAL Submitted (12-OCT-1993) J.E. Hodgson, Smithkline Beecham
Pharmaceuticals, Brockham Park, Betchworth, Surrey, RH3 7AJ, UK

FEATURES Location/Qualifiers

Source

1.4013
/organism="Staphylococcus aureus"
/strain="J2870"
/db_xref="taxon:1280"
/clone="POX 301"
1.450
/note="ORF B"
/codon_start=1
/transl_table=11
/protein_id="CA53190.1"
/db_xref="PID:9438227"
/db_xref="GI:438227"
/db_xref="SWISS-PROT:P41369"
/translation="EPEKHHVDLILNLTNSIGYKKDKLNLNLTSTSLKGGIILID
LNMPAFYKNFEQMLNFNEDYEVNQTINKLYNSIVSTWDYKNNEXSKDLSRELK
FKLYTNDITCSLYNFTLSNISFYGDMDKNGEHNPSRLVYKTK"

CDS

410..3588
/gene="iles"
410..415
/gene="1les"
434..439
/gene="1les"
461..466
/gene="1les"
/bound_molety="promoter"
477..3551
/gene="iles"
/codon_start=1
/transl_table=11
/product="isoleucyl tRNA synthetase"
/protein_id="CA53189.1"
/db_xref="PID:9581558"
/db_xref="GI:581558"
/db_xref="SWISS-PROT:P41368"
/translation="MTKKYLNTQNEISAFWNQKIPKSIDNKGQSEFVYDGPPTA
NGLPHAGVIGRYIKDLVARLKTMOGFYERKAGMDTGLPYVLEVEKKIGKQKIDF
EKYGIENFNECKSVENYEKWRDPSKDLGYVMDSPYITLNNYIESVNNLST
HKGLYKHKHKTPTCTHQTALSHSEVAGKKNVYDLSAVKFOITNSKDYTLSTWT
TTPWLPANVALINKDLNLSKIRVENEYIATDLSINSITKEKYIITDFSGSLNIN
LKIIPPESDGLNAVYVDEGVNTEGTYIATDILNAPAGEDYQVLRDPLNVI
TEGVYNDFPELVGNKAKNSDIEIKLSKOLYKKQKYEHNYPHCRGCGPLIY
AEGGFYKTTNFENKELIINNNTNEMFSPHKEBGMGNFLNEMNAGNRWGPEN
VVICNDNCHEVAPSSIKDLONNSINKIDDIELHRYVNYITLSCCKGKGRVVEY
IDVWPDGSGAPPAQHPTPDNOKIFNQHPPADITAGVDTKQWYSLVITLTKG
SSTKRALSGHILDSNGKMSKSGVNPTELINRYGADSLMALISDAPNNKRF
SENIVAQTSKFIIDLDNLYKEFNMYNKIDHYNPNEIKRSRTLDNMAISRLTLIK
ESNIVANNDETSAARLINEYNTISNMYIGDSRGPRWEGISNDKDAYNTLYEILT
TLSRIVAPVPFISEKIHNLGKSVHLDYDQYKESFNLQALDEMHVYIKVELSR
QARKNADIKIKOIKSMVYIKNSOLMSFLPNYSYTIKDELINKELINDNDIYTY
ELKLNSSVGRKLGNTKNIQITIDLSSEYDKSKLSIESNNFSLSDAELTKDDPITK
TLEKDSYQLSEDDCVILLDKNLSPELIRGARELIRLIQOARKKNLPINQRIIDY
IIVTGELLESIKNKNKMFENYIKNIHNLVDEYENTIFHNKEIKISILY"

CDS

3557..3588
/gene="1les"
3621..>4013
/note="ORF C"
/codon_start=1
/transl_table=11
/protein_id="CA53191.1"
/db_xref="PID:9438228"
/db_xref="GI:438228"
/db_xref="SWISS-PROT:P41370"
/translation="MYNFDYSKLPINKIQIKPIAGGYVLSFVSADSNKKYFLKLOP
NTKSNFDELSLKLTDKNIPVQIINKGELDNNSFLLEETENGHAYPESYRKLQ
KIYANMKNINSLNLFSGSHNFGTIEF"

CDS

stem_loop
1591 a 506 c 618 g 1298 t

BASE COUNT

ORIGIN

Query Match

12.6%; Score 40.4; DB 1; Length 4013;

Best Local Similarity 52.4%; Pred. No. 4.3;
Matches 86; Conservative 1; Mismatches 77; Indels 0; Gaps 0;

QY 13 TCTGTAGAAAATAGTACAGTTACACCAATGAGAACACAAAAAGTTGTGA 72
DB 3671 TATTTTTCGAATATTTTATGTTATGTTATGAAATATATACATAGTTATCA 3612
QY 73 CATTGTAGTAGGAGTGTCTACCCCTTACTCCCATCAAAAAAATGATACATGTT 132
DB 3611 CCTTTGCTTGCTTGTGATACATGACGACCTACTACTATAAATGATGTCATGTTGT 3552
QY 133 AAGGATATATAGGAGCAATATTTATCATATGTTCTTAAAGAGAA 176
DB 3551 TTAATATATATAGGAAATTTTATTTCTTATATATATATGAA 3508

RESULT 7
CEFA6C3
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS

REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS

REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS

REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS

true right end of clone F08812 is at 11562 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence 268104. The end of this sequence (29718..29777) overlaps with the start of sequence 268107.

FEATURES

source

Location/Qualifiers

1..29777
/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="X"

gene

/clone="F46C3"

complement(5728..7331)

CDS

/gene="F46C3.2"

complement(join(5728..5791,5836..6003,6061..6189,6339..6419,6467..6608,7046..7192,7242..7331))

/gene="F46C3.2"

/codon_start=1

/protein_id="CA91468.1"

/db_xref="PID:e1346869"

/db_xref="GI:3877185"

/db_xref="GI:3877185"

/db_xref="SPTREMBL:Q20457"

/translation="MWLVVGFQLPVKLTLLKGGMLRMQLTIMHVSVIDLYVTGM
DKRQKALNLTREGVLTDECFINISTHGVPERVQVPMRVYEMRIIDEGEDLOKE
KPKLKIETMKHYSASFNNHVRADLRSLFARATMPIMWTLFSPMTLAFASHI
MSOVNYSDLOETLSMTLPITLLVSSLLIGMLRYNYVMWGLCCENNESRLVTK
DDYSLLYHFCETLKEKFDRIYPRKGLQITLLMTYMLLVANHEHDLNITYPALVCVL
NGFGMLFARMSEFFRRCRTRDGTIAAR"

gene

complement(9727..26596)

CDS

/gene="hum-4"

/complement(join(9727..9807,10085..10186,10257..10423,10474..10616,11034..11173,11223..11383,11430..11604,11647..11780,11846..11979,12137..12416,12469..12577,12628..12776,12901..13082,13137..13261,13861..13879,14025..14261,14311..14431,14665..14638,14887..15013,15220..15302,15960..16008,16080..16201,16482..16676,16726..17201,17383..17609,17720..17873,17922..18046,18094..18288,18343..18477,18522..18682,18876..19118,19162..19930,19990..20160,20230..20339,20385..20525,20573..20737,20784..21343,21589..21783,21827..21993,23751..24069,25817..25911,26033..26348,26397..26596))
/note="Similarity to myosin: cDNA EST Yk249a4.5 comes from this gene: cDNA EST Yk270h6.5 comes from this gene: cDNA EST Yk284c3.5 comes from this gene: cDNA EST Yk353c4.5 comes from this gene: cDNA EST Yk470b4.5 comes from this gene: cDNA EST Yk284c3.3 comes from this gene: cDNA EST Yk470b4.3 comes from this gene"/codon_start=1
/protein_id="CA91469.1"
/db_xref="PID:e1346870"
/db_xref="PID:g3877186"
/db_xref="GI:3877186"
/db_xref="SPTREMBL:Q20456"/translation="MSTEVIALINPDHHTPDHTEANISRNKNSIACYNNMORTPIPL
NENSKAIFKISVKEBDLRLPIYDEPTAHPTMARNIYNPVSTFSRREGITA
VNLNCSLAKMRQHDHQLPNKLVTRICEDVRLDSNGNLSQSPFDVLEGLIVAN
HFLLEEFVPESETLLESETAEKFKKVPRLFCRSPNEVDSINLRICIDARVAV
FCYVMSNFTRLMISDRWEVETNOFILRWYDGMLELELLOSSTRAGVSE
EMESSTPIGFEYKETSKEPSMLASLILKPEKOPTKASVAVLVNPNKDSIDLSM
ETSTGTVANAVAKNALNKIFNNSSNAESTIVGEGSGSKSYNFAKFKYLLSPKRY
STRHSSIEFVRSFCACATLKNDENTRGCSIDLKRNVTGLMKTTPLELVRY
ISOKPERNNITIEVYHGLSDMKAKFGKQKFFYNQGNSSINIGHADVRFHL
ESALHVLGSDSDHCSYIKIISYILHIGNIGYKRPNPVEDVSSINIGHADVRFHL
LLEVDQVLFKLPSTSEGSTIELNALNDNSFAMLYEELFKVNLNIGLQVLS
LHTGVISILDHVGEFKYNNNGVEFLINSYNERIENLYKCHFDOLIVYAKGICSD
YKVPNSIENGKTYEILFKKPYGLISLLTDECKPCKGTHYELHCHNLNDRAYCA
RKNERLEPGRHICIGTIVNTPFARNKRIISLSAVOLMRNKNITIGLIPESYGN
TSOILVSOAFVLRGADIDAKLVNSHVHVRICISNNKOSTKEDIPVLRNRIKLL
LAEILLFRVKGYPVKISKTFARQYCLLPGLDIAOQNEKEIIODILQOGVAYEDF
KICTEVFLRERLAERYDIGNKICGDAIIIOKNNKSFVAQVYRKMRALIKLOSQ
LRGMKARBYIIRKEEMFKAIGRTMKRNRLDAYHVALGTEGSLQGLTVGIDINE

gene

CDS

DAKPLERPDSQDSEATETLHYLVPAKFNLSNMNSITLQFAEENFKHLEPRREP
IKTPLEHSDPDLVSIEIFKLLKYNKIDILTKROEDLGRYIVQOISPCORDE
ILVOTINIONKMDTASDNGKIVAHMISVPEPTEIIPMLIFPNNESVYMROLF
ATQRLRIKTIYSEIARELPSPSLIELAPINPVAEINCIDGLCCDHLHSMWTSSE
IAERILKRGVGLAGMTVEVETPNKVPAPGJNHITHVFSOIBGANDSEKQNTIF
NEPPEKLVKRPVPRVEVYDOBEIYVNTQNTNIPAAATYPIPEEMINQOPORRESN
TPRDKQKDEESDESLVREYENAGVATLLPRKNPAQAVEDNETSESEDEEDAN
RYGTEVYPERNNKMETIODESHILKSPVLPKRYISNDEOHEETPNAPPEPTY
POOMPMOYVPMKMTSMWTPSMTSPMPOQOILVMOQOMOPHESVYQYQPIQO
YCPPEPILSPQSVRESVPPMAKPVHDSGSTRSDYIIMKRGVPSYSTIRKMPV
EKGKDVDFLAVFDQVLSKQORAHNSNQLATINGKILRPGVITPEVQYISPV
PERPDLRQVDSPLRSKASLPRIISRHEEFYVPRSRNSNESTSDDONMYRT
RSERSLIPREHNGNYNDPSQVPMVQNGHEMELISPAESEAQSRRTHDR
SRHHVPGSVETMYAKRSVYDILKPPMSRRNDPMOTYRPLAOSPGROSRSF
EFSALPGDSARERMPNLLAROIIPRAYIPONKNYVPPSSYRASPSPTSGDR
GLNRLPQESYTPPAKNTKNNSEIISPRFNLDEOKAVIRHEKETAKANLNSRLK
AVKFAKAPKLTIRKEMFPGELNDIOLIDQVQIYEDCKSPYRAREQDRQVE
YLRQYEIIPSDLNQNSNHPVYKAVIELAMPLIYENQVYEVKRPDESSTIFA
ISEHGILRLVHTPHDLENPLKIDPEPPEITLADVLEANDILSVHREDEENAVSAV
RIKTNQAQIKRTILDRCLSGVPRKRYLATGSETPVGMKLYGKILENREGFLAQ
YVSDTDGVPPIRHEIETSDRDERVRFDEVPSESEYTMIDERTKTRFKDKKQ
ETWAMEDISQIVRESKPIQSOLADIGNESEKVAEYFHAIMKMGDEPLKSSMT
DVEFVYLICHRQPLRDVEYVQLIKQTSNSQKPNLSRAMRLITITIVQYFSSLT
LKPYVYLIGDADNEMOPHEVTCALICOTNMOTYKGRKYLNALAEQOITVGSV
TLYPCESTDGCOLRBOAFYISKDHVNSOILRPITVAFEMIOELCNLVNRSHEOE
SICLYVGDKNLHNTKRNNDYIMDITTESHKRLPPOFLIKRYVWHPLRYDPAIDS
MDOYIDIDLRLGSLISTNSLQOATATTEELIKLAVLPLLPDPKILNATLQIY
PSVIEPKRHRQEEVNTIRISRLQKMGGRMAEKASHLELTSTWPFEGVLAHRLKS
YVENGHQLEPVELITINKSGIOLQOPKSEVEFERNYDOIYVESIRKTAKYVRLVIN
TPQEGEIDIDIKDEADESHLIGQMFYTGAGEEGRTEL"
join(complement(3682..3795),complement(3460..3624),
complement(2750..3286),complement(2424..2700),
complement(2206..2376),complement(1899..2155),
complement(1748..1849),complement(1130..1425),
complement(709..1079),complement(185..659),
complement(105..304),complement(268104.1:26738..27119),
complement(268104.1:26574..26600))
/gene="F46C3.1"
join(complement(3682..3795),complement(3460..3624),
complement(2750..3286),complement(2424..2700),
complement(2206..2376),complement(1899..2155),
complement(1748..1849),complement(1130..1425),
complement(709..1079),complement(185..659),
complement(105..304),complement(268104.1:26738..27119),
complement(268104.1:26574..26600))
/note="similar to serine/threonine kinase (2 domains):
cDNA EST EMBL:D34010 comes from this gene: cDNA EST
EMBL:D3699 comes from this gene: cDNA EST EMBL:D64541
comes from this gene: cDNA EST EMBL:D67571 comes from this
gene: cDNA EST Yk459d11.3 comes from this gene: cDNA EST
Yk459d11.5 comes from this gene: cDNA EST Yk224a5.5 comes
from this gene: cDNA EST Yk251c3.5 comes from this gene:
cDNA EST Yk427c7.5 comes from this gene"
/codon_start=1
/protein_id="CA91470.1"
/db_xref="PID:e1346871"
/db_xref="PID:g3877187"
/db_xref="GI:3877187"
/db_xref="SPTREMBL:Q19192"

exon complement(2762, .2844)
 /note="MZEf, score = 60.4%"
 /evidence-not_experimental
 repeat_region 3260, .3470
 /rpt_family="LTR28"
 /evidence-not_experimental
 complement(3471, .3538)
 /rpt_family="LTR1A2"
 /evidence-not_experimental
 repeat_region 4631, .4700
 /note="homology = 70.00%, score = 21, counts = 10"
 /rpt_type="Lander"
 /rpt_unit="aaagaa"
 /evidence-not_experimental
 complement(4825, .4872)
 /note="GRAIL, score = 84.000%, comment = excellent"
 /evidence-not_experimental
 exon 5249, .5414
 /note="Genscan, score = 4.31%, comment = Terminal_exon 166
 bp frame: 2 phase: 1"
 /evidence-not_experimental
 repeat_region 6153, .6187
 /rpt_family="L2"
 /evidence-not_experimental
 complement(6701, .6701)
 /rpt_family="L2"
 /evidence-not_experimental
 repeat_region 6723, .7190
 /rpt_family="LTR1H"
 /evidence-not_experimental
 complement(6941, .6992)
 /note="Xpound exon prediction, score = 66% (0%)"
 /evidence-not_experimental
 repeat_region 7367, .7665
 /rpt_family="L2"
 /evidence-not_experimental
 complement(9762, .9962)
 /rpt_family="MER3"
 /evidence-not_experimental
 repeat_region 9964, .10150
 /rpt_family="MIR"
 /evidence-not_experimental
 complement(10362, .10639)
 /rpt_family="AluSg"
 /evidence-not_experimental
 complement(10363, .10619)
 /note="MZEf, score = 81.5%"
 /evidence-not_experimental
 complement(10895, .11136)
 /rpt_family="MIR"
 /evidence-not_experimental
 repeat_region 13416, .13515
 /rpt_family="MER5A"
 /evidence-not_experimental
 complement(14240, .14314)
 /rpt_family="MER5A"
 /evidence-not_experimental
 exon 14515, .14671
 /note="GRAIL, score = 95.000%, comment = excellent"
 /evidence-not_experimental
 complement(14515, .14671)
 /note="MZEf, score = 56.7%"
 /evidence-not_experimental
 repeat_region 14766, .14791
 /note="111 with 96% homology to 112"
 /rpt_type="inverted"
 /evidence-not_experimental
 complement(14808, .15551)
 /rpt_family="LTR2"
 /evidence-not_experimental
 complement(14957, .15105)
 /note="GRAIL, score = 63.000%, comment = good shadow"
 /evidence-not_experimental

exon complement(15431, .15537)
 /note="GRAIL, score = 47.000%, comment = marginal"
 /evidence-not_experimental
 repeat_region 15552, .15577
 /note="112 with 96% homology to 111"
 /rpt_type="inverted"
 /evidence-not_experimental
 repeat_region 15981, .17278
 /rpt_family="AluSg"
 /evidence-not_experimental
 complement(17569, .18083)
 /rpt_family="MER51B"
 /evidence-not_experimental
 complement(18364, .18887)
 /rpt_family="LIME2"
 /evidence-not_experimental
 complement(18884, .19040)
 /rpt_family="LIME2"
 /evidence-not_experimental
 complement(19041, .19170)
 /rpt_family="FLM_C"
 /evidence-not_experimental
 exon 19058, .19149
 /note="MZEf, score = 75.9%"
 /evidence-not_experimental
 complement(19171, .19315)
 /rpt_family="LIME2"
 /evidence-not_experimental
 complement(19314, .19857)
 /rpt_family="LIMEC"
 /evidence-not_experimental
 complement(19837, .19944)
 /rpt_family="LIMEC"
 /evidence-not_experimental
 repeat_region 20834, .20904
 /note="MZEf, score = 52.3%"
 /evidence-not_experimental
 complement(20834, .20904)
 /note="GRAIL, score = 70.000%, comment = good shadow"
 /evidence-not_experimental
 repeat_region 23833, .24258
 /rpt_family="LTR7"
 /evidence-not_experimental
 complement(24325, .24762)
 /rpt_family="LIP"
 /evidence-not_experimental
 repeat_region 24325, .24762
 /rpt_family="LIP"
 /evidence-not_experimental
 complement(24325, .24762)
 /note="MZEf, score = 87.1%"
 /evidence-not_experimental
 repeat_region 24763, .25071
 /rpt_family="AluJo"
 /evidence-not_experimental
 complement(25072, .25617)
 /rpt_family="LIP"
 /evidence-not_experimental
 complement(25096, .25247)
 /note="GRAIL, score = 71.000%, comment = good"
 /evidence-not_experimental

RESULT 12
AC007659/c
LOCUS
DEFINITION
AC007659
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC007659 89603 bp DNA 27-MAY-1999
Arabidopsis thaliana chromosome II BAC T14P1 genomic sequence,
complete sequence.
AC007659
94895233
AC007659.1 GI:4895233
HTG.
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurossids II; Brassicales; Brassicaceae;
Arabidopsis.
1 (bases 1 to 89603)
Goodman,H.M., Ayare,F., Belmonte,S., Dolan,M., Kim,U. and
Morris,U.W.
Arabidopsis thaliana chromosome II BAC T14P1 genomic sequence
Unpublished
2 (bases 1 to 89603)
Lin,X.
Direct Submission
Submitted (27-MAY-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
Address all correspondence to:
Howard Goodman
Mass. General Hospital
Boston MA 02114
goodman@hgh.harvard.edu

This BAC was annotated in collaboration with TIGR group.
Contact:
Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr.
Rockville, MD 20850, USA
e-mail: xlin@tigr.org
BAC clone T14P1 is from Arabidopsis chromosome II and is near the
molecular marker m36.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL (available by anonymous ftp
from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
Washington), Genscan (Chris Burge,
http://www.cbs.dtu.dk/netpgene/chsnetpgene.html), and NetPlantGene
(http://www.cbs.dtu.dk/netpgene/chsnetpgene.html), searches of the
complete sequence against a peptide database and the Arabidopsis
EST database at TIGR (http://www.tigr.org/tdb/at/at.html).
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by more
than two gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
genomic sequence that are not annotated as genes but have predicted
exons by GRAIL are annotated as misc features.

FEATURES
Source
Location/Qualifiers
1..89603
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"
/map="m36"
/clone="T14P1"
complement(1..539)
/gene="T14P1.1"

misc_feature
1..1868
/note="overlap with BAC clone F4L23
(AC002387:121061..122928)."
gene
complement(1..539)
/gene="T14P1.1"
/note="F4L23:31. identical to GB:AMB82643"
complement(97..501)
/gene="T14P1.1"
/note="unknown protein"
/codon_start=1
/protein_id="AAD32819.1"
/db_xref="PID:94895234"
/db_xref="GI:4895234"
/translation="MSKALAVTALITLNFEPFVTSKCPPTPKPKPKSPRK
APAVKPCPTDRLKIGVADLGIYVVGSPKPKPCCLLGLANLEAAVCLTRALK
AVYGINLVPIDLITLLVTCGRKVPHGQCS"
733..842
/note="exon predicted by xgrail, quality
marginal.shadowexon"
misc_feature
complement(791..852)
/note="exon predicted by xgrail, quality marginal"
complement(882..898)
/note="exon predicted by xgrail, quality marginal"
complement(join(1180..1571,1671..1723,2116..2171,
2294..2341,2499..2594))
/gene="T14P1.2"
complement(1180..2594)
/gene="T14P1.2"
/note="F4L23:32. identical to GB:AMB82646. Two ESTs do not
appear to be completely spliced"
complement(join(1352..1571,1671..1723,2116..2171,
2294..2333))
/gene="T14P1.2"
/codon_start=1
/product="putative symbiosis-related protein"
/protein_id="AAD32820.1"
/db_xref="PID:94895235"
/db_xref="GI:4895235"
/translation="NRKGSIFKDDDFEKRKAAGIRKRYPRIVYKAKSEYV
NIDKRLVYSDLTGQFVYVIRKRIKLSAEKATIFLVNVLPTGLMSVYEDKRD
EDGFLVYSGEFTGCASSI"
complement(3021..3042)
/rpt_family="AT-rich"
complement(3443..3571)
/note="exon predicted by xgrail, quality excellent"
3451..3563
/note="exon predicted by xgrail, quality
marginal.shadowexon"
misc_feature
complement(4034..4074)
/note="exon predicted by xgrail, quality good"
complement(4109..4140)
/note="exon predicted by xgrail, quality marginal"
4194..4302
/rpt_family="(TAAA)n"
complement(4657..4723)
/note="exon predicted by xgrail, quality good"
complement(5290..5327)
/rpt_family="(TA)n"
complement(5891..5921)
/note="exon predicted by xgrail, quality
marginal.shadowexon"
misc_feature
6344..6406
/note="exon predicted by xgrail, quality marginal"
complement(6361..6412)
/note="exon predicted by xgrail, quality
marginal.shadowexon"
repeat_region
6589..6645
/rpt_family="(GAAA)n"
complement(6722..6814)
/rpt_family="(GAAA)n"
complement(1..539)
/gene="T14P1.3"
<6845..8942
gene

FEATURES
source exons by GRALL are annotated as misc features.
Location/Qualifiers
1. 85534
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"
/map="m336"
/clone="T13E15"
join(<183. 313.385. 530.611. 775.902. >1020)
/gene="T13E15.1"
/note="predicted by genefinder"
join(<183. 313.385. 530.611. 775.902. 1020)
/gene="T13E15.1"
/note="hypothetical protein, 5' partial"
/codon_start=1
/protein_id="AAC31826.1"
/db_xref="PID:g2344886"
/db_xref="GI:2344886"
/translation="SKSYEPPIIPDDVDSGYNVYRGDCTNPTSYSHPAITFS
CVDSGNGGKGFDAISLSTNTVPATYHASEFKDLHLGDLIRIKIDDDQNTQA
SKPLWAVAVTQGSNRRKVPSSISIPDESCLASFSASQKSAIHPRIYQDYG
SDRSQMTVERLIRKRSYSITVKIFY"
join(<1966. 2016.2252. 2360.2553. 2799.2965. 3093,
3350. 3437.3518. 3805,3988. >4461)
/gene="T13E15.2"
/note="unknown protein"
/codon_start=1
/protein_id="AAC31827.1"
/db_xref="PID:g2344887"
/db_xref="GI:2344887"
/translation="MGEAGOSRDESEVTPLVGGNGDNNAVAPOVNSIPALNEASYI
TQATSYLSLEFGKDTISHPHEILSTSGVDNSPVSVCISGERPSTSEASTDA
NSPSRSTETILPATNAIVTSNRLNGLISMFGLEPARRIVRGSDIDGILQRPDE
MPVEDGTDFNKLIEDIGNHGPVYDITKTSKGLAKHAKIHSESSVENNAEII
KEYIEELCWGSKNRVLLIGHSGKIDAAALSLKXWELKRVGLVLAQSPYGGSPIA
TDLRGGQGDVNLARKMEIILSKYKGIQALDEDTYRRKEPLKNHLPRELPY
SERTEASIPAVYSLSHVAHAPLPLNQAKIPYMPICGAAMAACQILQVYXGKS
DELVTCDAEVPGSVYVRPKRLDHAMVTSISLNEVPLEDAAOVCEALITLVOVEQ
ERQOKLATKID"
complement(4471. 4553)
/rpt_family="(TAAA)n"
5103. 5185
/note="exon predicted by xgrail, quality marginal"
join(<5213. 5329,5510. 5671,5752. 5833,5953. 6180,
6279. >6481)
/gene="T13E15.3"
/note="predicted by grail"
join(5213. 5329,5510. 5671,5752. 5833,5953. 6180,
6279. 6481)
/gene="T13E15.3"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAC31828.1"
/db_xref="PID:g2344888"
/db_xref="GI:2344888"
/translation="MASTGEPRRRHRTSSISPSAAAVKOPFEWPSSEDLDAV
LQFONLKLQLEAQVECSILEKLSQIEKLPYSSSKTYHRSKQTLASVESC
SVASDSSGAAHVRNKEDGSPAVKNDIFNRLLETGATSSSNICSNMENGAVT
SSOMOTIVNVAATEDILKIDELVPTVIRNGIDICOLALSELESIKRFPRLD
DVLVFKSLRELSQHRDADAVRDLKIRSYLVSSSTIYC"
6935. 6996
/note="exon predicted by xgrail, quality marginal"
complement(7199. 7378)
misc_feature
misc_feature

mRNA
/note="exon predicted by xgrail, quality good_shadowexon"
join(<7745. 7778,7833. 8051,8142. 8302,8385. 8456,
8670. 8746,9120. 9314,9426. 9498,9591. 9677,9916. 10053,
10139. 10341,10436. 10601,10727. >10858)
/gene="T13E15.4"
/note="unknown protein"
/codon_start=1
/protein_id="AAC31829.1"
/db_xref="PID:g2344889"
/db_xref="GI:2344889"
/translation="NALERKLVISYKRSIRKQVERDSIWKEREININELGDV
SKRTSAVDSRNASLDSEIQLQDEKMRITKLGINSREGRKEIFADKALITSEPE
EKSSMRSQLNRYKETAGIGHSIRADVOSISGVLCRKTEYEAQSDYASQDLN
ATYCDLKNSEHEIKLEIDYKRESRDVARDIAKADQYRAHVAQSLKSLDQNL
RYKAANEAAVAGSQMAAFAEADLRQMDCKRDVAKHSDILKSKHEHETYS
QTIGSAEDIVPQNOQLQYTERDPTNKLLEGTSROMQDTLLIDKINDKIOQ
GSATVSLKSSRLDQRFCTDQFQRLAEKTKRSYLENDQKRRADIGLEQAR
SLRESHSVQESRLDYGALDELETERENRRIEEMEARKRVSRSLLEGSSAI
GNLRQELSEFEIKLCKACNDPRKEVYITKCYHLFCNCPQVRLTGTROKCPCSASF
GNDIRPIYI"
11224. 11350
/note="exon predicted by xgrail, quality marginal"
complement(11229. 11346)
/note="exon predicted by xgrail, quality good_shadowexon"
complement(11574. 11658)
/note="exon predicted by xgrail, quality marginal_shadowexon"
11605. 11675
/note="exon predicted by xgrail, quality marginal"
11707. 11717
/note="exon predicted by xgrail, quality good"
12218. 12317
/rpt_family="(TAAA)n"
complement(12427. 12530)
/rpt_family="(TAAA)n"
complement(12865. 13011)
/note="exon predicted by xgrail, quality marginal"
12886. 12968
/note="exon predicted by xgrail, quality good_shadowexon"
13215. 13246
/rpt_family="AT_rich"
complement(13891. 13930)
/rpt_family="AT_rich"
14229. 14256
/rpt_family="AT_rich"
complement(14814. 14894)
/rpt_family="AT_rich"
complement(<15279. 16184)
/gene="T13E15.5"
complement(<15279. 16184)
/gene="T13E15.5"
complement(15279. 16166)
/gene="T13E15.5"
/codon_start=1
/product="TINY transcription factor 1solog"
Query Match 12.18; Score 38.8; DB 8; Length 85534;
Best Local Similarity 48.58; Pred. No. 7.2; Mismatches 0; Gaps 0;
Matches 100; Conservative 2; Indels 0;
QY 33 GTTACAACCAATGGAACAACAAGAAATTGTCATGTCATGAGGAGTGTG 92
1118 GTTGGATCATACACAATAATGACAAAAGATTTAATGAAAAATCAATTCGG 11129
QY 93 ACCCCTTACTCCCAATCAAAAAAATGATACATGCTTAAGATARAAGGCATAT 152
Db 11128 AAACCATGTCCAAAATTAACAAGAAAGAGTATCACCAGACATAACCTTAACG 11069

11 CTCTCTGTAGAAAATAGTACAGTTACACCAATAGGAACACACAAAAGAAAAGTTTGT 70

```

Db 65351 CTTCGATTTATTAACGACATTTTAACATGATGAAATGAAAAAACTGACGCAATGTGG 65410
Oy 71 GACATGTAGTAGGGAGCTGTGTACCCCTTACTCCCATCAAAAAAAATGATACATCG 130
Db 65411 TATGAGAGAGTCAACATGAAATTTCTCTAGGGAGCTGAAAAAATGATATGACATGT 65470
Oy 131 TTAAGCATATAAAGGCGCAATTTTATCATATGTCATTAAGAGAGAGAGAAAAATAC 190
Db 65471 ATTAACATTTAGGACGACAGAAATATTTTACTCTATTTCAATTTTAAAGATGACATATGTCC 65530
Oy 191 TACTTCTCRAAATGGAAGCCCTTAA 216
Db 65531 TATGCTCTTAAGATGAGCAATTA 65556

RESULT 15
HS800F24/c 138888 bp DNA PRI 12-MAY-1999
LOCUS Human DNA sequence from clone 800F24 on chromosome 1q24, complete
DEFINITION
sequence.
ACCESSION 299758
NID 94775628
VERSION 299758.7 GI:4775628
KEYWORDS HMG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 138888)
AUTHORS Grafham,D.
TITLE Direct Submission
JOURNAL Submitted (12-MAY-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
O On May 11, 1999 this sequence version replaced q1:4678515.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL
IMPORTANT: This sequence is not the entire insert of clone 800F24.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone d11018K9 is at 138789 in this sequence.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 1, constructed by the Sanger Centre Chromosome 1
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
800F24 is from the library RPC14 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://dnapac.med.buffalo.edu/VECTOR: pcYPAC2.
location/Qualifiers
1. 138888
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="800F24"
/map="q24"
/clone_id="RPC14"
1. 1964
/note="11 repeat: matches 3306..5335 of consensus"
repeat_region
1967..2267

```

```
/note-"AluSc repeat: matches 1. .300 of consensus"
repeat_region 2272. .2340
/note-"LIP2 repeat: matches 5413. .5484 of consensus"
repeat_region 2340. .3242
/note-"LIM1 repeat: matches 799. .1710 of consensus"
repeat_region 3351. .3836
/note-"LIM1 repeat: matches 491. .794 of consensus"
repeat_region 4306. .4426
/note-"MIR repeat: matches 87. .216 of consensus"
repeat_region 5580. .5677
/note-"MIR repeat: matches 49. .146 of consensus"
repeat_region 7200. .7639
/note-"LIR16B repeat: matches 1. .464 of consensus"
repeat_region 8511. .8829
/note-"AluSq repeat: matches 1. .309 of consensus"
repeat_region 10092. .10203
/note-"L2 repeat: matches 2613. .2710 of consensus"
repeat_region 11207. .11257
/note-"MIR repeat: matches 200. .250 of consensus"
repeat_region 11301. .11504
/note-"AluIb repeat: matches 85. .287 of consensus"
repeat_region 11664. .11708
/note-"MIR repeat: matches 72. .116 of consensus"
repeat_region 13586. .13665
/note-"40 copies 2 mer aa 64% conserved"
repeat_region 13784. .14056
/note-"L2 repeat: matches 1484. .1801 of consensus"
repeat_region 15463. .15706
/note-"LIMEC repeat: matches 2371. .2296 of consensus"
repeat_region 15739. .16233
/note-"LIR1D repeat: matches 1. .500 of consensus"
repeat_region 16772. .17099
/note-"LIM1 repeat: matches 3067. .3032 of consensus"
repeat_region 17100. .17412
/note-"AluSx repeat: matches 1. .312 of consensus"
repeat_region 17413. .17859
/note-"LIM1 repeat: matches 2633. .3068 of consensus"
repeat_region 17897. .18009
/note-"LIPAB repeat: matches 6050. .6162 of consensus"
repeat_region 18077. .18220
/note-"LIM4 repeat: matches 2352. .2163 of consensus"
repeat_region 18345. .18494
/note-"75 copies 2 mer tt 60% conserved"
repeat_region 18506. .18646
/note-"LIMEC repeat: matches 1851. .1992 of consensus"
repeat_region 18654. .19000
/note-"LIM4C repeat: matches 1057. .1405 of consensus"
repeat_region 19414. .19509
/note-"48 copies 2 mer tt 63% conserved"
repeat_region 20280. .20351
/note-"LIR1 repeat: matches 91. .141 of consensus"
repeat_region 20516. .20618
/note-"LIR1H repeat: matches 429. .543 of consensus"
repeat_region 21226. .21528
/note-"AluSx repeat: matches 1. .305 of consensus"
repeat_region 21802. .21932
/note-"MIR repeat: matches 14. .148 of consensus"
repeat_region 23153. .23265
/note-"MER41B repeat: matches 258. .370 of consensus"
repeat_region 24524. .24628
/note-"MIR repeat: matches 18. .141 of consensus"
repeat_region 24629. .24927
/note-"AluSx repeat: matches 1. .299 of consensus"
repeat_region 24928. .24986
/note-"MIR repeat: matches 141. .194 of consensus"
repeat_region 25802. .25928
/note-"FLAM_C repeat: matches 2. .131 of consensus"
repeat_region 25939. .26052
/note-"L2 repeat: matches 1594. .1714 of consensus"
repeat_region 26056. .26416
/note-"L2 repeat: matches 1890. .2200 of consensus"
repeat_region 26733. .27086
/note-"LIR40b repeat: matches 86. .446 of consensus"

repeat_region 27202. .27345
/note-"L2 repeat: matches 2559. .2710 of consensus"
repeat_region 27849. .28052
/note-"MIR repeat: matches 1. .205 of consensus"
repeat_region 28914. .29215
/note-"AluSx repeat: matches 1. .310 of consensus"
repeat_region 29892. .30582
/note-"LIPAB repeat: matches 5476. .6155 of consensus"
repeat_region 30596. .31979
/note-"LIPAB repeat: matches 4744. .6155 of consensus"
repeat_region 32709. .32764
/note-"L2 repeat: matches 2654. .2709 of consensus"
repeat_region 34717. .35145
/note-"L2 repeat: matches 1906. .2371 of consensus"
repeat_region 35142. .35391
/note-"L2 repeat: matches 2141. .2359 of consensus"
repeat_region 35392. .35691
/note-"AluSx repeat: matches 1. .300 of consensus"
repeat_region 35692. .37115
/note-"L2 repeat: matches 477. .2141 of consensus"
repeat_region 37415. .37828
/note-"MSTB repeat: matches 6. .426 of consensus"
repeat_region 38428. .38572
/note-"MIR repeat: matches 111. .262 of consensus"
repeat_region 38626. .39246
/note-"LIM4 repeat: matches 4297. .4933 of consensus"
repeat_region 39346. .39651
/note-"AluIO repeat: matches 1. .301 of consensus"
repeat_region 39699. .40087
/note-"L2 repeat: matches 1879. .2298 of consensus"
repeat_region 40158. .40665
/note-"LIM4 repeat: matches 4796. .5332 of consensus"
repeat_region 40666. .40992
/note-"TIGER1 repeat: matches 2078. .2418 of consensus"
repeat_region 41012. .42354
/note-"LIPAB repeat: matches 4784. .6144 of consensus"
repeat_region 43109. .43233
/note-"MIR repeat: matches 20. .148 of consensus"
repeat_region 43333. .43694
/note-"L2 repeat: matches 1188. .1584 of consensus"
repeat_region 44121. .44349
/note-"MER8 repeat: matches 8. .239 of consensus"
repeat_region 44360. .44465
/note-"L2 repeat: matches 2602. .2712 of consensus"
repeat_region 44502. .44695
/note-"L2 repeat: matches 2302. .2491 of consensus"
repeat_region 45695. .45961
/note-"AluIO repeat: matches 1. .299 of consensus"
repeat_region 46075. .46116
/note-"21 copies 2 mer at 79% conserved"
repeat_region 46137. .46200
/note-"32 copies 2 mer tt 67% conserved"
repeat_region 46646. .46790
/note-"LIME repeat: matches 5607. .5751 of consensus"
repeat_region 46899. .47049
/note-"MIR repeat: matches 102. .258 of consensus"
repeat_region 47326. .47575
/note-"MER46A repeat: matches 1. .236 of consensus"
repeat_region 47845. .48314
/note-"LIP repeat: matches 4916. .5395 of consensus"
repeat_region 48303. .52827
/note-"LIPAB repeat: matches 1236. .5781 of consensus"
repeat_region 52826. .57156
/note-"LIPAB repeat: matches 2. .4330 of consensus"
repeat_region 57157. .57376
/note-"AluY repeat: matches 94. .311 of consensus"
repeat_region 57377. .59193
/note-"LIPAB repeat: matches 4320. .6144 of consensus"
repeat_region 58567. .58667
/note-"Single clone region"
repeat_region 59220. .59596

Query Match 12.0%; Score 38.4; DB 10; Length 138888;
```


This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 12:30:02 ; Search time 289.74 seconds
(without alignments)
276.322 Million cell updates/sec

Title: US-09-030-606-224

Perfect score: 320

Sequence: 1 CCCCTGAGGCTTCTGTGTA.....TTTARACTCMGATGTGAC 320

Scoring table: IDENTITY_NNC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	318	99.4	320	V58694	Prostate tumour sp
2	318	99.4	320	V61269	CDNA sequence of p
3	45	14.1	110000	V21209_01	Continuation (2 of
4	37	11.6	110000	V21209_15	Continuation (16 o
5	36.6	11.4	19124	T72882	Plasmidum var-7 g
6	35.6	11.1	1215	T68041	H. pylori cytoplas
7	35.6	11.1	717	T67711	H. pylori cytoplas
8	35	10.9	1107	V74949	Staphylococcus aut
9	34.6	10.8	1811	V17183	Insecticidal toxin
10	34.2	10.7	10813	V74575	Staphylococcus aut
11	34.2	10.7	110000	X20248_04	Continuation (5 of
12	34.2	10.7	116277	X20249	Borrelia burgdorfe
13	34	10.6	9842	X20259	Borrelia burgdorfe
14	33.8	10.6	110000	X20248_03	Continuation (4 of
15	33.2	10.4	7295	T72883	Plasmidum var-1 g
16	32.8	10.2	64976	V21209_16	Continuation (17 o
17	32.8	10.2	2120	X33812	Coding sequence fo
18	32.6	10.2	110000	X20248_08	Continuation (9 of
19	32.4	10.1	1406	T41619	Tomato endo-xylogl
20	32.4	10.1	14736	V52304	Streptococcus pneu
21	32.2	10.1	400	V87584	EST clone DY307. N
22	32	10.0	4093	O49264	Med-4. Long-distan
23	32	10.0	1457	V15409	Methicillin resist
24	32	10.0	80595	V83939	HC-contig derived
25	32	10.0	80240	V83940	NC-contig derived
26	32	10.0	18359	X20255	Borrelia burgdorfe
27	31.8	9.9	1920	O88066	Bacillus lichenifo
28	31.8	9.9	1920	O95031	Bacillus lichenifo
29	31.8	9.9	1920	V02471	Bacillus lichenifo
30	31.8	9.9	110000	V21209_10	Continuation (11 o
31	31.6	9.9	134525	O04525	Total base sequenc
32	31.6	9.9	5852	O11710	Dicystostellum plas
33	31.6	9.9	5596	V83941	Bacterial artifical
34	31.4	9.8	2207	N80748	Sequence of a poly
35	31.4	9.8	1753	V18455	Human granulocytic
36	31.4	9.8	110000	X20248_01	Continuation (2 of
37	31.4	9.8	110000	X20248_02	Continuation (3 of
38	31.4	9.8	10715	X20248_09	Continuation (10 o
39	31.2	9.8	4472	T05406	Human cell cycle g
40	31	9.7	4059	T08973	CI gene. Transgent
41	31	9.7	110000	V21209_13	Continuation (14 o
42	31	9.7	110000	V21209_14	Continuation (15 o
43	31	9.7	2025	V54207	Nucleotide sequenc

ALIGNMENTS

44 31 9.7 3072 1 V54163
c 45 31 9.7 4059 1 X15630

Nucleotide sequenc
Nucleotide sequenc

RESULT 1

V58694
ID V58694 standard; CDNA: 320 BP.
AC V58694;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone p5105.
KW Prostate tumour specific gene; human; prostate cancer; detection; therapy; ss.
OS Homo sapiens.
PN W09837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J.
DR MPI; 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful for detecting and treating prostate cancers
PS Claim 1, Page 134, 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can be used in the method of the invention. The method is for detecting prostate cancer comprises contacting a biological sample with an agent able to bind an immunogenic portion of a prostate protein (such as CC encoded by this sequence). An antibody which binds to an immunogenic portion of the prostate protein, and the method can be used to detect, monitor progression of, or treat prostate cancers. The antibody may CC also be conjugated to a therapeutic agent for use in therapy of prostate cancers.
SQ Sequence 320 BP; 111 A; 56 C; 64 G; 84 T;

Query Match 99.4%; Score 318; DB 1; Length 320;

Best Local Similarity 100.0%; Pred. No. 1.3e-74;

Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCTGAGGCTTCTGTGTAAGAAATAGTACAGTACCAACATAGAACACAAAAA 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1 CCCCTGAGGCTTCTGTGTAAGAAATAGTACAGTACCAACATAGAACACAAAAA 60
DB
QY 61 AAAAGTTTGACATGTGTAGGAGAGTGATCCCTTACCTCCCATCAAAAAAAAT 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
61 AAAAGTTTGACATGTGTAGGAGAGTGATCCCTTACCTCCCATCAAAAAAAAT 120
DB
QY 121 GGATACATGTTTAAAGATARAAGGCAATATTTATCATATGTTCTAAAGAGAA 180
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
121 GGATACATGTTTAAAGATARAAGGCAATATTTATCATATGTTCTAAAGAGAA 180
DB
QY 181 GAAAAAATACATCTTCTCAAAATGAGAGCCCTTAAAGTGCTTTATACAGAGAC 240
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
181 GAAAAAATACATCTTCTCAAAATGAGAGCCCTTAAAGTGCTTTATACAGAGAC 240
DB
QY 241 AAATGAGCGGCTCATCTCCCTTAAAGTGATGATGATGATGATGATGATGATGAT 300
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
241 AAATGAGCGGCTCATCTCCCTTAAAGTGATGATGATGATGATGATGATGATGAT 300
DB
QY 301 TTTARACTCMGATGTGAC 320
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
301 TTTARACTCMGATGTGAC 320
DB
RESULT 2
ID V61269 standard; CDNA: 320 BP.
AC V61269;

06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P510S.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN W0837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; 003492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI) CORIAX CORP.
PI Dillon DC, Xu J.
DR WPI: 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PS used in a vaccine for the treatment of prostate cancer
PT Claim 3; Page 124; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 320 BP; 111 A; 56 C; 64 G; 84 T;

Query Match 99.4%; Score 318; DB 1; Length 320;
Best Local Similarity 100.0%; Pred. No. 1.3e-74;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCCCTGAAGGCTTCTGTTAGAAAATAGTACAGTTACAAACCAATGAGAACAAACAAAAGA 60
DB 1 CCCCTGAAGGCTTCTGTTAGAAAATAGTACAGTTACAAACCAATGAGAACAAACAAAAGA 60
OY 61 AAAATGTTGTGACATGTGTAGTGGAGTGTCACCCCTTACCCCTCAATGAGAACAAACAAAAGA 120
DB 61 AAAATGTTGTGACATGTGTAGTGGAGTGTCACCCCTTACCCCTCAATGAGAACAAACAAAAGA 120
OY 121 GGATACATGTTAAAGATARAAGGCAATATTTATCATATGTTCTAAAGAGAA 180
DB 121 GGATACATGTTAAAGATARAAGGCAATATTTATCATATGTTCTAAAGAGAA 180
OY 181 GAGAAAATCTACTTCTCRAAATGAGACCCCTTAAAGTGCTTTGANAAGAGACAC 240
DB 181 GAGAAAATCTACTTCTCRAAATGAGACCCCTTAAAGTGCTTTGANAAGAGACAC 240
OY 241 AATGTGGCCGTCACCTTAAAGTGTGATGATGCTGGACAGGTACGTTGCACT 300
DB 241 AATGTGGCCGTCACCTTAAAGTGTGATGATGCTGGACAGGTACGTTGCACT 300
OY 301 TTARACTCMGATGTGAC 320
DB 301 TTARACTCMGATGTGAC 320

RESULT 3
V21209_01/c
Continuation (2 of 17) of V21209 from base 100001 (Methanococcus jannaschii circular chi
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

WP	Fragment Name	Begin	End
WP	V21209_00	1	110000
WP	V21209_01	100001	210000
WP	V21209_02	200001	310000
WP	V21209_03	300001	410000
WP	V21209_04	400001	510000
WP	V21209_05	500001	610000
WP	V21209_06	600001	710000
WP	V21209_07	700001	810000
WP	V21209_08	800001	910000
WP	V21209_09	900001	1010000
WP	V21209_10	1000001	1110000
WP	V21209_11	1100001	1210000
WP	V21209_12	1200001	1310000
WP	V21209_13	1300001	1410000

WP V21209_14 1400001 1510000
WP V21209_15 1500001 1610000
WP V21209_16 1600001 1664976

Query Match 14.1%; Score 45; DB 1; Length 110000;
Best Local Similarity 48.2%; Pred. No. 0.0093;
Matches 120; Conservative 2; Mismatches 127; Indels 0; Gaps 0;

OY 5 TGAAGGCTTCTGTTAGAAAATAGTACAGTTACAAACCAATGAGAACAAACAAAAGA 64
DB 93590 TGAAGACCACTTAAAGATGAGAAATATTTAAACCAATATCCAAATATATTAATA 93531
OY 65 GTTGTGACATGTGTAGTGGAGTGTCACCCCTTACCCCTCAATGAGAACAAACAAAAGA 124
DB 93530 ATTGTGTAAGTTTAAAGACAGATGAGAAATAGTACCTTCCAAAACAAAATGTT 93471
OY 125 ACATGTTAAAGATARAAGGCAATATTTATCATATGTTCTAAAGAGAAAGAGA 184
DB 93470 TAAATGGACAGCATGACAAACTAATTTTATAAATAATCTAATTAATAAAGCAAT 93411
OY 185 AATATCTACTTCTCRAAATGAGACCCCTTAAAGTGCTTTGATGATGAGACACAAAT 244
DB 93410 AATATGTTCCAACTTGAAGACAAACCAATATATAGTTTGTAGGAGATGATGAAT 93351
OY 245 GTGGCCGTC 253
DB 93350 ATTACCTTC 93342

RESULT 4
V21209_15/c
Continuation (16 of 17) of V21209 from base 1500001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

WP	Fragment Name	Begin	End
WP	V21209_00	1	110000
WP	V21209_01	100001	210000
WP	V21209_02	200001	310000
WP	V21209_03	300001	410000
WP	V21209_04	400001	510000
WP	V21209_05	500001	610000
WP	V21209_06	600001	710000
WP	V21209_07	700001	810000
WP	V21209_08	800001	910000
WP	V21209_09	900001	1010000
WP	V21209_10	1000001	1110000
WP	V21209_11	1100001	1210000
WP	V21209_12	1200001	1310000
WP	V21209_13	1300001	1410000
WP	V21209_14	1400001	1510000
WP	V21209_15	1500001	1610000
WP	V21209_16	1600001	1664976

RESULT	ID	Accession	Standard	CDNA	Length
727882	DT	T72882	standard	19124 BP	
DT	12-SEP-1997	(first entry)			
DE	Plasmodium var-7 gene.				
KW	DBL gene family; SAMP, stalic acid binding protein; vaccine; therapy;				
KW	Duffy binding like gene; Duffy antigen binding protein; erythrocyte;				
KW	DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;				
OS	Plasmodium vivax.				
OS	Plasmodium falciparum.				
FT	Key	Location/Qualifiers			
FT	exon	7317..15139			
FT	intron	15140..16205			
FT	exon	16206..17552			
FT	intron	17553..18611			
FT	exon	18612..19124			
FT	intron	19125..20183			
FT	exon	20184..21242			
FT	intron	21243..22301			
FT	exon	22302..23360			
FT	intron	23361..24419			
FT	exon	24420..25478			
FT	intron	25479..26537			
FT	exon	26538..27596			
FT	intron	27597..28655			
FT	exon	28656..29714			
FT	intron	29715..30773			
FT	exon	30774..31832			
FT	intron	31833..32891			
FT	exon	32892..33950			
FT	intron	33951..35009			
FT	exon	35010..36068			
FT	intron	36069..37127			
FT	exon	37128..38186			
FT	intron	38187..39245			
FT	exon	39246..40304			
FT	intron	40305..41363			
FT	exon	41364..42422			
FT	intron	42423..43481			
FT	exon	43482..44540			
FT	intron	44541..45599			
FT	exon	45600..46658			
FT	intron	46659..47717			
FT	exon	47718..48776			
FT	intron	48777..49835			
FT	exon	49836..50894			
FT	intron	50895..51953			
FT	exon	51954..53012			
FT	intron	53013..54071			
FT	exon	54072..55130			
FT	intron	55131..56189			
FT	exon	56190..57248			
FT	intron	57249..58307			
FT	exon	58308..59366			
FT	intron	59367..60425			
FT	exon	60426..61484			
FT	intron	61485..62543			
FT	exon	62544..63602			
FT	intron	63603..64661			
FT	exon	64662..65720			
FT	intron	65721..66779			
FT	exon	66780..67838			
FT	intron	67839..68897			
FT	exon	68898..70000			
FT	intron	70001..71059			
FT	exon	71060..72118			
FT	intron	72119..73177			
FT	exon	73178..74236			
FT	intron	74237..75295			
FT	exon	75296..76354			
FT	intron	76355..77413			
FT	exon	77414..78472			
FT	intron	78473..79531			
FT	exon	79532..80590			
FT	intron	80591..81649			
FT	exon	81650..82708			
FT	intron	82709..83767			

Db	522	AATATAATTAACTAGGAAGAGTAAACATATTTT	564
RESULT	6		
ID	T68041		
AC	T68041 standard; DNA; 1215 BP.		
NC	T68041:		
DE	H. pylori cytoplasmic protein ORF 07gp11807orf12.		
KW	Cytoplasmic; vaccine; prevention; treatment; infection; identification;		
RK	binding compound; bacterium; life cycle; activator; bacteria; inhibitor;		
OS	duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.		
EH	Helicobacter pylori.		
FT	Key	Location/Qualifiers	
FT	cds	1..1215	
FT	/*tag= a		
FN	/note= "no stop codon given"		
PD	M09640893-AI.		
PR	19-DEC-1996.		
PR	06-JUN-1996; U09122.		
PR	07-JUN-1995; US-487032.		
PR	01-APR-1996; US-630405.		
PA	(ASTR) ASTRA AB.		
PI	Berglindh OT, Smith D, Mellgaard BJ,		
DR	WPI, 97-052306/05.		
P-PSDB:	W20788.		
PT	Helicobacter pylori nucleic acid sequences and related		
PT	polypeptide(s) - useful for vaccines to treat or prevent H. pylori		
PT	infection, and to detect Helicobacter		
PS	Claim 9: Page 862: 1481bp. English.		
CC	This sequence encodes a H. pylori cytoplasmic protein.		
CC	The protein may be used in a vaccine to prevent or treat H. pylori		
CC	infection or to identify H. pylori polypeptide binding compounds,		
CC	useful as potential H. pylori life cycle activators or inhibitors.		
CC	The genomic sequence of H. pylori (ATCC 55679) was determined from		
CC	overlapping contigs generated by mechanically shearing the bacterial		
CC	DNA. The sequences were analysed for ORF of at least 180 nucleotides,		
CC	and the predicted coding regions defined by computer evaluation. To		
CC	identify likely H. pylori antigens for vaccine development, the amino		
CC	acid sequences predicted from various ORF were analysed for significant		
CC	homology to other known or exported membrane proteins. Having identified		
CC	and determined the sequences of interest, particular regions can be		
CC	isolated from H. pylori by PCR amplification for recombinant polypeptide		
SC	production, e.g. in E. coli hosts.		
Sequence	1215 BP; 379 A; 254 C; 214 G; 368 T;		
Query Match	11.1%; Score 35.6; DB 1; Length 1215;		
Best Local Similarity	56.0%; Pident. No. 0.93;		
Matches	65; Conservative 1; Mismatches 50; Indels 0; Gaps 0.		
OY	128 TGGTAAAGATARAAGGCAATATTTATCATGTTCCTAAGAAGAAAGAAA 187		
	:		
DB	857 TGGGTAAGTGCGCAACCGCTGATTTATTTGTTAGATTTTTTAAAGAAATGAAAGAAAAT 916		
OY	188 TACTACTTCTCRAATGGAAGGCCCTTAAGAGCTGCTTGATACCTGAGACACAAA 243		
DB	917 TCATCATTTTTTCAAGCGGCAACCGCTTTAAACCCTTAAACCGCATGCCTTAA 972		
RESULT	7		
ID	T67711		
AC	T67711 standard; DNA; 717 BP.		
NC	T67711:		
DE	H. pylori cytoplasmic protein ORF 719606.aa.		
KW	Cytoplasmic; vaccine; prevention; treatment; infection; identification;		
RK	binding compound; bacterium; life cycle; activator; bacteria; inhibitor;		
OS	duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.		
EH	Helicobacter pylori.		
FT	Key	Location/Qualifiers	
FT	cds	1..717	

```

FT      /*tag= a
PN      /note= "no stop codon given"
PD      MO9640893-A1.
PR      19-DEC-1996.
PR      06-JUN-1996; U09122.
PR      07-JUN-1995; US-487032.
PR      01-APR-1996; US-630405.
PA      (ASNR ) ASTRA AB.
PI      Berglindh OT, Smith D, Mellgaard BJ.
PI      WPI: 97-052306/05.
DR      P-PSDB: W20567.
PR      Helicobacter pylori nucleic acid sequences and related
PR      polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PR      infection, and to detect Helicobacter
PS      Claim 9; Page 315-318; 1481pp; English.
CC      This sequence encodes a H. pylori cytoplasmic protein.
CC      The protein may be used in a vaccine to prevent or treat H. pylori
CC      infection or to identify H. pylori polypeptide binding compounds,
CC      useful as potential H. pylori life cycle activators or inhibitors.
CC      The genomic sequence of H. pylori (ATCC 55679) was determined from
CC      overlapping contigs generated by mechanically shearing the bacterial
CC      DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC      and the predicted coding regions defined by computer evaluation. To
CC      identify likely H. pylori antigens for vaccine development, the amino
CC      acid sequences predicted from various ORF were analysed for significant
CC      homology to other known or exported membrane proteins. Having identified
CC      and determined the sequences of interest, particular regions can be
CC      isolated from H. pylori by PCR amplification for recombinant polypeptide
CC      production, e.g. in E. coli hosts.
SQ      Sequence 717 BP; 224 A; 127 C; 146 G; 220 T;

Query Match
Best Local Similarity 11.1%; Score 35.6; DB 1; Length 717;
Matches 65; Conservative 1; Mismatches 50; Indels 0; Gaps 0;

OY      128 TGGTTAAAGATARAAGGCAATATTATCATATGTTCTAAAGAGAAGAGAAA 187
DB      434 TGGGTAATGCGCCAGCCTGATTTATGTTGATTTTAAAGAAATGAAGAAAAT 493
OY      188 TACTACTTTCRARAATGGAAGCCCTTAAGAGTCTTGATATCTGAGACACAAA 243
DB      494 TCATCATTTTTCATGCGGAGACCGCTTTAAAGCCTTAAAGCGCATGGCTTAA 549

RESULT 8
V74949 8
ID      V74949 standard; DNA: 1107 BP.
AC      V74949;
DT      16-MAR-1999 (first entry)
DE      Staphylococcus aureus contig SEQ ID #638.
KW      Computer readable medium; vaccine; S. aureus infection; immunodetection;
KW      cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW      skin infection; surgical wound infection; scalded skin syndrome;
KW      toxic shock syndrome; ds.
OS      Staphylococcus aureus.
PA      Key Location/Qualifiers
PI      misc_feature 541..600
PI      /*tag= a
PI      /note= "these bases represent a line of missing text in
PI      the sequence listing in the specification. They
PI      are included to maintain the nucleotide numbering
PI      given in the specification for this DNA sequence"

EP-786519-A2.
PD      30-JUN-1997.
PR      07-JAN-1997; 100117.
PR      05-JAN-1996; US-009861.
PA      (HUMA-) HUMAN GENOME SCI INC.
PI      Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI      Rosen CA;
PI      WPI: 97-374922/35.
DR      Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT      stored on computer readable medium and used in the production of

```

```

PR      anti-S. aureus vaccines
PS      Claim 1; Page 1565-1566; 3271pp; English.
CC      This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC      of the invention. The DNA sequences are recorded on a computer readable
CC      medium, preferably selected from a floppy or hard disk, random access
CC      memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC      the S. aureus DNA sequences allows putative functions to be assigned so
CC      that protein-encoding or regulatory regions of commercial, therapeutic or
CC      industrial importance can be obtained. Specifically, sequences which are
CC      likely to encode antigens have been identified and these polypeptides can
CC      be used in a vaccine composition against S. aureus infection. The
CC      polypeptides can also be used in a kit for the immunodetection of
CC      S. aureus in a sample. S. aureus is implicated in numerous human diseases,
CC      including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC      skin and surgical wound infections, scalded skin syndrome, toxic shock
CC      syndrome, etc. Organisms transformed with the DNA sequences can be used
CC      for recombinant production of the polypeptides. The new DNA sequences
CC      (and their fragments) are useful as primers or probes for isolating
CC      homologues of any of the S. aureus DNA sequences contained on the
CC      computer readable medium.
SQ      Sequence 1107 BP; 384 A; 134 C; 194 G; 329 T;

Query Match
Best Local Similarity 10.9%; Score 35; DB 1; Length 1107;
Matches 56; Conservative 2; Mismatches 39; Indels 0; Gaps 0;

OY      108 TCAAAAAAATGATACATGTTAAAGATARAAGGCAATATTATCATATGTTCT 167
DB      221 TAAAAATAAAAATGCGGAGGATTAATGCTAAACATCAAGAGCTTACATCTTCT 280
OY      168 AAAAGAGAAGAGAAATACATCTACTTCTCRAAT 204
DB      281 AAAAAGGACAAAGCATATACCGTTTCGTAGAT 317

RESULT 9
V7183 9
ID      V7183 standard; DNA: 1811 BP.
AC      V7183;
DT      18-JUN-1998 (first entry)
DE      Insecticidal toxin subunit BrHTX-1(a)1.1 genomic sequence.
KW      Insecticidal toxin; Bracon hebetor; insect control; pathogen;
KW      recombinant baculovirus; ds.
OS      Bracon hebetor.
PA      Key Location/Qualifiers
PI      W09744355-A1.
PI      27-NOV-1997.
PI      PD 01-MAY-1997; G01205.
PI      PR 22-MAY-1996; GB-010748.
PI      PR 22-MAY-1996; GB-010687.
PI      PR 22-MAY-1996; GB-010695.
PI      PR 22-MAY-1996; GB-010697.
PI      PR 22-MAY-1996; GB-010738.
PI      PR 22-MAY-1996; GB-010739.
PI      PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PI      PA (ZENE ) ZENECA LTD.
PI      PI Christian PD, Claudianos A, Daly A, Duncan RE, Smer M,
PI      Windass JD,
PI      WPI: 98-018430/02.
DR      New nucleic acid encoding a combination of insecticidal sub:unit(s)
PT      of wasp toxin - and related transformed cells, insect pathogens and
PT      combinations of proteins, useful as insecticides
PS      Claim 1; Pages 57-58; 84pp; English.
CC      This is a contiguous genomic sequence which encodes part of a spliced RNA
CC      that is related to the cDNA comprised by pBrHTX-1(a)1.1. This 1811 base
CC      pair spliced RNA derived from the Bracon hebetor genomic clone encodes
CC      at least two of the insecticidal toxin subunits shown in sequences
CC      W52124-W52128. The spliced RNA can hybridise with extension products
CC      prepared from a 564 base pair (V71745) template with 6 specified primers.
CC      A nucleic acid encoding at least one of the specified subunits can be
CC      modified so that mRNA instability motifs and/or fortuitous splice sites
CC      are removed, or insect-pest preference codons are used, so that
CC      expression of this nucleic acid in insect cells yields practically the

```

WP	X20248_09	900001	910715
----	-----------	--------	--------

Query Match 10.7%; Score 34.2; DB 1; Length 110000;
Best Local Similarity 47.8%; Pred. No. 6.1; Mismatches 104; Indels 0; Gaps 0;
Matches 96; Conservative 1; Mismatches 104; Indels 0; Gaps 0;

QY 12 TTTCTGTGAAATAGTACAGTTACACCAATAGACACAAAGAAAGTTGTG 71
DB 39340 TCGTTGAAATACATATTTAAATAAGAAACAAAGAAAGTTGAAACCTCG 39399
QY 72 ACATTGTAGTAGGAGTGTGACCCCTTCTCTCCCATCAAAAAAATGATACATGCT 131
DB 39400 CTAATGTAGCAATTAATTTTATTTTATTTTATTAATAAAGTCCAGCATTTAG 39459
QY 132 TAAAGATARAAGGCAATATTTTATCATATGTTCTAAAGAAAGAGAGAAATACT 191
DB 39460 TAAATTAATCTAGTGTGATTTTATTTTATTTTATTTTATTTTAAATTAAGCAAAATTTAA 39519
QY 192 ACTTCTCRAAATGGAAGCC 212
DB 39520 AATATCTCAAAATTTAACTC 39540

RESULT 12
X20249/c
AC X20249 standard; DNA; 116277 BP.
AC X20249:
AT 04-MAY-1999 (first entry)
DE Borrelia burgdorferi polynucleotide sequence #2.
KM Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KM epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KM infection; diagnosis; characterisation; detection; ds.
OS Borrelia burgdorferi.
PN WO9858943-A1.
PD 30-DEC-1998.
PF 18-JUN-1998; U12764.
PR 03-SEP-1997; US-057483.
PR 20-JUN-1997; US-050359.
PR 22-JUL-1997; US-053344.
PR 22-JUL-1997; US-053377.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI White OR;
PI WPI: 99-081217/07.
DR New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
PS Claim 1: Page 672-737; 1128pp; English.
CC X20248 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease
SQ Sequence 116277 BP; 42656 A; 19868 C; 14490 G; 39250 T;

Query Match 10.7%; Score 34.2; DB 1; Length 116277;
Best Local Similarity 47.8%; Pred. No. 6.2; Mismatches 104; Indels 0; Gaps 0;
Matches 96; Conservative 1; Mismatches 104; Indels 0; Gaps 0;

QY 12 TTTCTGTGAAATAGTACAGTTACACCAATAGACACAAAGAAAGTTGTG 71
DB 115794 TCGTTGAAATACATATTTAAATAAGAAACAAAGAAAGTTGAAACCTCG 115735
QY 72 ACATTGTAGTAGGAGTGTGACCCCTTCTCTCCCATCAAAAAAATGATACATGCT 131
DB 115734 CTAATGTAGCAATTAATTTTATTTTATTTTATTTTATTTTAAATTAAGTCCAGCATTTAG 115675

QY 132 TAAAGATARAAGGCAATATTTTATCATATGTTCTAAAGAAAGAGAAATACT 191
DB 115674 TAAATTAATCTAGTGTGATTTTATTTTATTTTATTTTATTTTAAATTAAGCAAAATTTAA 115615
QY 192 ACTTCTCRAAATGGAAGCC 212
DB 115614 AATATCTCAAAATTTAACTC 115594

RESULT 13
X20259/c
ID X20259 standard; DNA; 9842 BP.
AC X20259:
AT 04-MAY-1999 (first entry)
DE Borrelia burgdorferi polynucleotide sequence #12.
KM Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KM epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KM infection; diagnosis; characterisation; detection; ds.
OS Borrelia burgdorferi.
PN WO9858943-A1.
PD 30-DEC-1998.
PF 18-JUN-1998; U12764.
PR 03-SEP-1997; US-057483.
PR 20-JUN-1997; US-050359.
PR 22-JUL-1997; US-053344.
PR 22-JUL-1997; US-053377.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI White OR;
PI WPI: 99-081217/07.
DR New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
PS Claim 1: Page 914-919; 1128pp; English.
CC X20248 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease
SQ Sequence 9842 BP; 3671 A; 1321 C; 1151 G; 3699 T;

Query Match 10.6%; Score 34; DB 1; Length 9842;
Best Local Similarity 46.6%; Pred. No. 3.9; Mismatches 108; Indels 0; Gaps 0;
Matches 97; Conservative 3; Mismatches 108; Indels 0; Gaps 0;

QY 99 TACTCCCATCAAAAAAATGATACATGTTAAAGATARAAGGCAATATTTATC 158
DB 6937 TATTTGGCAATATTAATTAATAAATTTACTTCATAAAAAACATATGATCATATTTATT 6878
QY 159 ATATGTTCTAAAGAAAGAAAGAAATACTACTTCTCRAAATGGAAGCCCTTAAG 218
DB 6877 ACATATTTAAAGATTAATAAGATATATATTTGATTAATCTGAACCTTAATAAACATTTACT 6818
QY 219 GTGCTTGATGATGAAGACACAAATGTGGCGTCCATCTCTTARAGTGCATGACT 278
DB 6817 AAAAAAGCAAAAAAACACACTTTAGGAATCCGATACCGGGCGGAGATGCAAT 6758
QY 279 TGGACACGTTAATCTTGCAGTTTARA 306
DB 6757 TGTAGATGATCCTGATGGCAGATAGAA 6730

RESULT 14
X20248-03
Continuation (4 of 10) of X20248 from base 300001 (Borrelia burgdorferi polynucleotide
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248
WP Fragment Name Begin End

WP X20248_00 1 110000
 WP X20248_01 100001 210000
 WP X20248_02 200001 310000
 WP X20248_03 300001 410000
 WP X20248_04 400001 510000
 WP X20248_05 500001 610000
 WP X20248_06 600001 710000
 WP X20248_07 700001 810000
 WP X20248_08 800001 910000
 WP X20248_09 900001 910715

Query Match 10.6%; Score 33.8; DB 1; Length 110000;
 Best Local Similarity 53.4%; Pred. No. 7.8;
 Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

OY 42 AATAGGACACACAAAAGTTGTGACATGTAGAGGAGTGTACCCCTTAC 101
 DB 14419 AATATCTCAACATATATAAAGAACTATAGCAATATAATGCTTAAATATCTCCCAA 14478
 OY 102 TCCCATTCACAAAAAATGATGATCATGTTAAGGATATAGGCAATTTTATCATTA 161
 DB 14479 ACTTCAACATGATATATAATATATATGTAAGCTTATATATCCCATAGCCACATTT 14538

OY 162 TGTTCTTAAAGAG 174
 DB 14539 CTTTCAAAAATAG 14551

RESULT 15

T72883
 AC T72883 standard; cDNA; 7295 BP.
 DT 12-SEP-1997 (first entry)
 DE Plasmodium var-1 gene.
 KW DBL gene family; SABP; stalic acid binding protein; vaccine; therapy;
 KM Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 OS Plasmodium vivax.
 OS Plasmodium falciparum.
 FH key Location/Qualifiers
 FT cds 517..7065
 FT /tag- a
 FT /product- var-1
 PN MO9640766-A2.
 PD 19-DEC-1996.
 PF 07-JUN-1996; 009508.
 PR 07-JUN-1995; US-487826.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Chitniss C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 DR WPI: 97-052231/05.
 DR P-PSDB: W22476.
 PT New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and stalic acid
 PT binding proteins
 PS Claim 4; Page 67-69; 96pp; English.
 CC This sequence represents the var-1 gene of Plasmodium. Var-1 belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and stalic acid binding protein
 CC (SABP) conserved regions (see T72889 and T72888 respectively). The var
 CC family of genes modulate cytoadherence and antigenic variation of
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 CC protein (DABP) are soluble proteins that appear in the culture
 CC supernatant after infected erythrocytes release merozoites. DABP and SABP
 CC mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the invention.
 CC The compositions are for the treatment and prevention of malaria, and
 CC comprise either a nucleotide sequence or encoded polypeptide of the
 CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
 CC genes having homology with conserved regions of DABP and SABP. The

CC compositions are used for the treatment and prevention of malaria. They
 CC are also used in the preparation of vaccines for inducing a protective
 CC immune response in a mammal to Plasmodium merozoites (especially
 CC Plasmodium falciparum or Plasmodium vivax).
 SQ Sequence 7295 BP; 2866 A; 1067 C; 1455 G; 1907 T;

Query Match 10.4%; Score 33.2; DB 1; Length 7295;
 Best Local Similarity 51.4%; Pred. No. 5.9;
 Matches 74; Conservative 1; Mismatches 69; Indels 0; Gaps 0;

OY 15 TTGTAGAAAAATAGTACAGTACACCAATAGCAACACAAAAAGTTGTGACA 74
 DB 4885 TGCGTAGAATATTTTACAGATATATATAGAAATAGAAAAATTAAGCTATGTACG 4944
 OY 75 TTGTAGTAGGAGTGTGTACCCCTTACTCCCATCAAAAAAATGATATATGTTAA 134
 DB 4945 AAAAAGGAGATGATGCAATGTATATTAAGTTGTATGAAAAAATGGCTACAGAAAA 5004
 OY 135 AGGATATAAGGCAATATTTATC 158
 DB 5005 ACGAAGAAATGCAAAAAATTAAC 5028

Search completed: September 28, 1999, 12:30:20
 Job time: 5440 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:34:15 ; Search time 161.06 Seconds
(without alignments)
182.146 Million cell updates/sec

Title: US-09-030-606-224

Perfect score: 320

Sequence: 1 CCCCTGAGAGCTTCTGTGTA.....TTTARACTCGCATGTGAC 320

Scoring table: IDENTITY_NUC

Searched: 176461 seqs, 45838279 residues

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/5C.COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/5D.COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PTUS9.COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38.2	11.9	7218	2	US-08-232-463-14
2	34.4	10.8	2367	3	US-08-441-139-3
3	31.8	9.9	1930	3	US-08-720-899-1
4	31.8	9.9	1930	3	US-08-459-610-1
5	31.8	9.9	1920	4	US-08-343-804-1
6	31.6	9.9	5852	1	US-07-867-106-2
7	31	9.7	4059	4	US-08-485-139-1
8	30.6	9.6	4435	2	US-08-308-872B-5
9	30.6	9.6	331	4	US-08-520-678A-23
10	30.4	9.5	9636	3	US-08-323-170B-1
11	30.2	9.4	19557	5	PCT-US92-06300-1
12	29.8	9.3	12124	1	US-08-181-271A-36
13	29.8	9.3	12124	2	US-08-449-803-36
14	29.8	9.3	12124	2	US-08-444-803-36
15	29.8	9.3	12124	2	US-08-449-043-36
16	29.8	9.3	12124	2	US-08-456-263A-36
17	29.8	9.3	12124	3	US-08-455-416-36
18	29.8	9.3	12124	3	US-08-455-244-36
19	29.8	9.3	12124	3	US-08-454-876-36
20	29.8	9.3	12124	4	US-08-457-364-36
21	29.8	9.3	12124	4	US-08-456-262-36
22	29.8	9.3	12124	4	US-08-456-240-36
23	29.8	9.3	12124	4	US-08-455-736-36
24	29.6	9.2	15108	1	US-08-157-005-1
25	29.6	9.2	15101	5	PCT-US95-09927-14
26	29.4	9.2	1834	1	US-08-347-792-5
27	29.4	9.2	1824	2	US-08-431-357-5
28	29.4	9.2	2277	3	US-08-676-967-2
29	29.4	9.2	2291	3	US-08-355-888A-6
30	29.4	9.2	2277	3	US-08-676-974-2
31	29.4	9.2	2606	4	US-08-568-459A-7
32	29.4	9.2	2991	4	US-08-588-190-1
33	29.4	9.2	2891	4	US-08-693-697-6
34	29.4	9.2	2880	4	US-08-693-697-32
35	29.4	9.2	2877	4	US-08-693-697-35
36	29.4	9.2	246240	4	US-08-724-394A-20
37	29.4	9.2	246240	4	US-08-724-394A-21

38	29.4	9.2	246240	4	US-08-724-394A-22	Sequence 22, Appl
39	29.4	9.2	2880	4	US-08-588-526-1	Sequence 1, Appl
40	29.4	9.2	1824	5	PCT-US95-15353-5	Sequence 5, Appl
41	29	9.1	5181	3	US-08-257-073-10	Sequence 10, Appl
42	29	9.1	11832	4	US-08-416-603-1	Sequence 1, Appl
43	29	9.1	356	4	US-08-520-678A-22	Sequence 22, Appl
44	29	9.1	257	4	US-08-520-678A-24	Sequence 24, Appl
45	29	9.1	253	4	US-08-520-678A-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 diagonal Road, Suite 500
; City: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)836-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fis
; US-08-232-463-14

Query Match 11.9%; Score 38.2; DB 2; Length 7218;
Best Local Similarity 8.7%; Pred. No. 0.033;
Matches 24; Conservative 138; Mismatches 113; Indels 0; Gaps 0;

QY 20 AGAATAAGTACGTTTACCAACCATAGAACACAAAAAGTTTGACATTGTA 79
DB 1289 RR 1330
QY 80 GTAGGAGTGTGATCCCTACTCCCATCAAAAAAATGATCATGTTAAAGAT 139

[illegible]

```

RESULT 2
US-08-441-139-3
: Sequence 3, Application US/08441139
: Patent No. 5773245
: GENERAL INFORMATION:
: APPLICANT: Wilttrup, Dr. Karl D.
: APPLICANT: Robinson, Anne S.
: TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
: TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
: STREET: 400 Garden City Plaza
: CITY: Garden City
: STATE: NY
: COUNTRY: USA
: ZIP: 11530
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/441,139
: FILING DATE: 15-MAY-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/089,997
: FILING DATE: 06-JUL-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Digiglio, Frank S.
: REGISTRATION NUMBER: 31,346
: REFERENCE/DOCKET NUMBER: 8646
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 516-742-4343
: TELEFAX: 516-742-4366
: TELEX: 230 901 SANS UR
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2367 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 251..2176
: US-08-441-139-3

```

	Query Match	Similarity	Score	DB	Length
Best Local	50	Conservative	64.1%	Pred. No. 0, 27	Mismatches 27; Indels 0; Gaps 0
QY	112	AAAAAAAAATGATACATGCTTAAAGAGATRAAAGGCCATATTTATCATATGCTTAAAA	171		
Db	2275	AAAGAAAGATGCTGGATCTTTTGAAGAGATTAAGTATCATATATTTTAAATGCTTAAAA	2334		

QY 172 GAGAAGGAGAGAAATA 189
+ + + + + + + + +
Db 2335 AAAAAAAAAAAAAAAAAA 2352

```

RESULT 3
US-08-720-899-1
Sequence 1, Application US/08720899
Patent No. 5753460
GENERAL INFORMATION:
APPLICANT: Bisgaard-Frantzen, Henrik
APPLICANT: Bisgaard, Torben Vedel
APPLICANT: Svendsen, Allan
APPLICANT: Thellersten, Marianne
APPLICANT: Van der Zee, Pia
TITLE OF INVENTION: AMYLASE VARIANTS
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5753460 No. 5753460disk of No. 5753460th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/720, 899
FILING DATE: 10-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/343, 804
FILING DATE: 22-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lowmney Dr., Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4054.214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 334..1872
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 334..420
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 421..1869
US-08-720-899-1

```

	Query Match	Similarity	Score	DB	Length
Best Local	54	9.9%	31.8	3	1920
Matches	54	Conservative	2	Mismatches	39
				Indels	0
				Gaps	0
QY	110	AAAAAAAAATGATATACATGCTTAAAGCATAAAGGCGCAATTTTATCATATGTTCTAA	169		
DB	187	AGAACGAGAGAGCGCTATTGCAATTAATGATTAAGCGCCATATCGGCGCTTTCTTTGG	246		
QY	170	AAGAGAGAGAGAGAAATACTACTCTTCRAAT	204		
DB	247	AAGAAATATATAGGAAATGCTACTCTGTTAAAAAT	281		

```

RESULT 4
US-08-459-610-1
: Sequence 1, Application US/08459610
: Patent No. 5801043
: GENERAL INFORMATION:
: APPLICANT: Bisgaard-Frantzen, Henrik
: APPLICANT: Borchert, Torben Vedel
: APPLICANT: Svendsen, Allan
: APPLICANT: Thellersen, Marianne
: APPLICANT: Van der Zee, Pia
: TITLE OF INVENTION: AMYLASE VARIANTS
: NUMBER OF SEQUENCES: 38
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 58010430 No. 5801043disk of No. 5801043th America, Inc.
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/459,610
: FILING DATE: 02-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/343,804
: FILING DATE: 22-NOV-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Lorney Dr., Karen A.
: REGISTRATION NUMBER: 31,274
: REFERENCE/DOCKET NUMBER: 4054, 214-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1920 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 334..1872
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 334..420
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: 421..1869
: US-08-459-610-1

Query Match 9.9%; Score 31.8; DB 3; Length 1920;
Best Local Similarity 56.8%; Pred. No. 1.4;
Matches 54; Conservative 2; Mismatches 39; Indels 0; Gaps 0

QY 110 AAAAAAATGATACATGTTAAAGCATARAAGGCAATATTTATCATATGTTCTAA 169
Db 187 AGAAGCAGAGAGGATGATTAATAAATGATAGAGCCCATATGCGCGCTTTCTTTGG 246

QY 170 AAGAGAGGAGAGAGAAATACACTCTTCRCRAAT 204
Db 247 AAGAAAAATATAGGAGAAAATGCTACTTGTAAAAAT 281

RESULT 5
US-08-343-804-1
: Sequence 1, Application US/08343804

```

```

Patent No. 5830837
GENERAL INFORMATION:
APPLICANT: Bisgaard-Frantzen, Henrik
APPLICANT: Borchert, Torden Vedel
APPLICANT: Svendsen, Allan
APPLICANT: Thellersen, Marianne
APPLICANT: Van der Zee, Pia
TITLE OF INVENTION: AMYLASE VARIANTS
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSSEE: NO. 58308370 No. 5830837disk of No. 5830837th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentlin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,804
FILING DATE: 22-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lowney Dr., Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4054, 214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 334..1872
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 334..420
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 421..1869
US-08-343-804-1

Query Match          9.9%; Score 31.8; DB 4; Length 1920;
Best Local Similarity 56.8%; Pred. No. 1.4;
Matches 54; Conservative 2; Mismatches 39; Indels 0; Gaps 0;

OY 110 AAAAAAAAAATGCATACATGCTTAAAGCATARAAGCGCAATTTTATCATATGTCTCAA 169
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 187 AGAAGCAGAAGAGCTATGATAAATGAATGATGAAGCAGCATATTCGGCGTTTCTTTGG 246
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 170 AAGAGAGAGAGAGAAAATACTACTCTTCTCRAAT 204
      ||||| | | |||||| | |||| | : |||
DB 247 AAGAAAAATATAGGAAAATGTACTTGTTAAAAAT 281
      ||||| | | |||||| | |||| | : |||

RESULT        6
US-07-867-106-2
Sequence 2, Application US/07867106
Patent No. 5389526
GENERAL INFORMATION:
APPLICANT: Slade, Martin B
APPLICANT: Chang, Andy C M
APPLICANT: Williams, Keith L
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
                    Slime Moulds of the Genus Dictyostelium

```

```

Query Match          9.9%; Score 31.6; DB 1; Length 5852;
Best Local Similarity 51.5%; Pred. No. 2.4;
Matches 70; Conservative 1; Mismatches 65; Indels 0; Gaps 0;

OY      36 ACAACCAATAGAGAACAAAGAAAAGTTGGACATTGTAGTAGGAGATGTGTACC 95
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5449 AAAAAAAAAAAAAAAAAAAAAAAAAAATCAAAATGTCTTTGTTTAAAGATTTTAAATC 5508

OY      96 CCTTACCTCCCATCAAAAAAATATGATACATGGTTAAAGGTAAAGGCAATATTTT 155
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5509 TCGTCAATGATTTTAAATATAAATCGATACATTAATTTAAAAAACCCTTTACATTTT 5568

OY      156 ATCATATGTGTTCTAATA 171
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5569 TTATTTTAATTCCTCAA 5584

RESULT 7
US-08-485-139-1/c
: Sequence 1, Application US/08485139
: Patent No. 5880331
: GENERAL INFORMATION:
: APPLICANT: KREBBERS, ENNO
: APPLICANT: WILLIAMS, MARK
: APPLICANT: LEEMANS, JAN
: TITLE OF INVENTION: USE OF ANTHOCYANIN GENES TO MAINTAIN
: TITLE OF INVENTION: MALE STERILE PLANTS
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
:

```

```

NAME/KEY: -
LOCATION: 1505..1510
OTHER INFORMATION: /label= ECORI
FEATURE:
NAME/KEY: -
LOCATION: 2081..2086
OTHER INFORMATION: /label= XhoI
FEATURE:
NAME/KEY: -
LOCATION: 2418..2430
OTHER INFORMATION: /label= SfiI
FEATURE:
NAME/KEY: -
LOCATION: 2669..2674
OTHER INFORMATION: /label= SnaBI
FEATURE:
NAME/KEY: -
LOCATION: 2634..2639
OTHER INFORMATION: /label= SnaBI
FEATURE:
NAME/KEY: -
LOCATION: 3008..3013
OTHER INFORMATION: /label= HpaI
FEATURE:
NAME/KEY: -
LOCATION: 1..1077
OTHER INFORMATION: /label= PciI
OTHER INFORMATION: /note= "region containing promoter of C1 gene"
FEATURE:

```

```
NAME/KEY: LOCATION: 1078..2134
OTHER INFORMATION: /label- C1
OTHER INFORMATION: /note- "coding region of C1 gene"
FEATURE:
NAME/KEY: LOCATION: 2135..2430
OTHER INFORMATION: /label- 3'C1
OTHER INFORMATION: /note- "region containing polyadenylation signal"
OTHER INFORMATION: of C1 gene
FEATURE:
NAME/KEY: LOCATION: 1033..1038
OTHER INFORMATION: /label- TATA-Box
FEATURE:
NAME/KEY: LOCATION: 1061..1062
OTHER INFORMATION: /label- transcript-init
OTHER INFORMATION: /note- "transcription initiation site"
FEATURE:
NAME/KEY: LOCATION: 1211..1299
FEATURE:
NAME/KEY: LOCATION: 1430..1575
FEATURE:
NAME/KEY: LOCATION: 935..939
OTHER INFORMATION: /label- C1-S
OTHER INFORMATION: /note- "TGCAG sequence (in C1 gene) which in the
OTHER INFORMATION: C1-S sequence is changed to TAGG"
US-08-485-139-1

Query Match          9.78; Score 31; DB 4; Length 4059;
Best Local Similarity 61.78; Pred. No. 3.1;
Matches 66; Conservative 0; Mismatches 40; Indels 1; Gaps 1;

CY 18 TTAAATAATATACAGTACACCAATGAGAACAAACAAAAGTTGTGACATG 77
DB 3636 TTAAATAATGACACAAATTCATCATATAGACACAA-ATAAGAGAAAGTTGTGACATCA 3578

CY 78 TAGTAGGAGGTGTACCCCTTACTCCCAATCAAAAAAATGAT 124
DB 3577 AGTCTAAATAAGTCCCATGAGTCAACTATATAAAAAAGATGAT 3531

RESULT 8
US-08-308-872B-5/C
; Sequence 5, Application US/08308872B
; Patent No. 5661006
; GENERAL INFORMATION:
; APPLICANT: BROWN, Thomas David Kay
; APPLICANT: HORSBURGH, Brian Collin
; TITLE OF INVENTION: CANINE CORONAVIRUS SUBUNIT VACCINE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5661006el Patent Department
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/308.872B
; FILING DATE: 19-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042.846
```

```
FILING DATE: 05-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,641
FILING DATE: 24-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 91.303.737.0
FILING DATE: 25-APR-1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gortmley, Mary E.
REGISTRATION NUMBER: 34,409
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4435 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Canine corona virus
STRAIN: CCV-V54
FEATURE:
NAME/KEY: CDS
LOCATION: 60..4418
OTHER INFORMATION: /label- CCV-C54_Spikegene
US-08-308-872B-5

Query Match          9.68; Score 30.6; DB 2; Length 4435;
Best Local Similarity 47.58; Pred. No. 4.1;
Matches 87; Conservative 1; Mismatches 95; Indels 0; Gaps 0;

CY 2 CCCTGAAGGCTTCTTTAGAAATAGTACAGTTACCAACCAATGAGAACAAAGAA 61
DB 429 CTCTATAGGTGATGATGTAATTAACATGATTAACAGAGCATGACATGACTGTA 370

CY 62 AAGATTGTGACATGTAGTAGGAGGTGTACCCCTTACTCCCATCAAAAAAATG 121
DB 369 AAGTTTACCTCTTGCATTCGCCAGTACTATTCGCCATGCTTCATATCAAAATTAATG 310

CY 122 GATACATGTTAAAGGTAATAGGCAATATTTATCATATGTTCTAAAGAGAGAG 181
DB 309 CAGTATGTTACTAAATTAATGTAAGCTGATGTTGCTGCTTGAACAGATTGACC 250

CY 182 AGA 184
DB 249 ACA 247

RESULT 9
US-08-520-678A-23/C
; Sequence 23, Application US/08520678A
; Patent No. 5874565
; GENERAL INFORMATION:
; APPLICANT: Rice, Charles M.
; APPLICANT: Kolykhalov, Alexander A.
; TITLE OF INVENTION: NOVEL 3' TERMINAL SEQUENCE OF HEPATITIS
; TITLE OF INVENTION: C VIRUS GENOME AND DIAGNOSTIC AND THERAPEUTIC USES THEREOF
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howell & Haferkamp, L.C.
; STREET: 7733 Forsyth Blvd., Suite 1400
; CITY: St. Louis
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
```

```

LENGTH 9636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 149..9556
US-08-323-170B-1

Query Match          9.5%; Score 30.4; DB 3; Length 9636;
Best Local Similarity 52.5%; Pred. No.6.2;
Matches 64; Conservative 1; Mismatches 57; Indels 0; Gaps 0

QY 108 TCAAAAAAAAAATGATATCATGTGTTAAAGATARAAGGCAATATTTATCATATAGTTCT 167
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 4379 TCAAAAAACAAATGTGGAAAAATCATATTAGAGAGAAAAATATATATCATATTACTGTTCA 4438

QY 168 AAAAGAGAGAGAGAGAGAAATTAATCTCTTCTTCRAAATGAGACCCCTTAAGGTGCTTTA 227
      |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 4439 AAAAGAGAGAGAGAGAAAGTAAAGGTGTTCTATTGTGATAATGAACATGCACATATGTTTACT 4498

QY 228 TA 229
      ||
Db 4499 TA 4500

RESULT 11
PCT-US92-06300-1
: Sequence 1, Application PC/TUS9206300
: GENERAL INFORMATION:
: APPLICANT: Hurwitz, David R
: APPLICANT: Nathan, Margret
: APPLICANT: Shanl, Moshe
: TITLE OF INVENTION: Transgenic Protein Production
: NUMBER OF SEQUENCES: 1
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Rhone-Poulenc Rorer, Inc.
: STREET: 500 Virginia Ave., Bldg. 3A
: CITY: Ft. Washington
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19034
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/06300
: FILING DATE: 19920730
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: Goodman, Rosanne
: REGISTRATION NUMBER: 32,534
: REFERENCE/DOCKET NUMBER: A0856-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 962-4130
: TELEFAX: (215) 962-4107
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH 19557 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: PUBLICATION INFORMATION:
: AUTHORS: Minghetti, P P
: AUTHORS: Ruffner, D E
: AUTHORS: Kuang, W.-J.

```


QY 49 ACACACAAAAAGAAATGTTGTGCACATTGTAGTAGGGAGTGTGTACCCCTTACTCCCCAT 108
||| ||||| ||||| ||| | ||||
Db 10697 ACAGCAAAAAAAAAAATTCGACAGGTTGTGATCATTGTCAATGATGTCAATAATATAT 10756
||| ||||| ||||| ||| | ||||
QY 109 CAAAAAAAATGCGATCATGTGTTTAAGAGATARAAGGCCAATATTTTATCATATGTTCTGA 168

APPLICANT: Moyet, Mary B.
 APPLICANT: Neuhaus, Jean-Marc
 APPLICANT: Payne, George B.
 APPLICANT: Sperlison, Christoph
 APPLICANT: Stinson, Jeffrey R.
 APPLICANT: Uknes, Scott J.
 APPLICANT: Ward, Eric R.
 APPLICANT: Williams, Sherlica C.
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
 TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: New York
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/444,803
 FILING DATE: 19-MAY-1995
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/181,271
 FILING DATE: 13-JAN-94
 APPLICATION NUMBER: US 08/093,301
 FILING DATE: 16-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/937,197
 FILING DATE: 6-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/678,378
 FILING DATE: 1-APR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/305,566
 FILING DATE: 6-FEB-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/165,667
 FILING DATE: 8-MAR-1988
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/042,847
 FILING DATE: 6-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/632,441
 FILING DATE: 21-DEC-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/425,504
 FILING DATE: 20-OCT-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/848,506
 FILING DATE: 6-MAR-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/768,122
 FILING DATE: 27-SEP-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/580,431
 FILING DATE: 7-SEP-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/368,672
 FILING DATE: 20-JUN-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/329,018
 FILING DATE: 24-MAR-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/045,957
 FILING DATE: 12-APR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmer, James Scott

```

1 REGISTRATION NUMBER: 36,129
2 REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
3 TELECOMMUNICATION INFORMATION:
4 TELEPHONE: (919)541-8614
5 TELEFAX: (919)541-8689
6 INFORMATION FOR SEQ ID NO: 36:
7 SEQUENCE CHARACTERISTICS:
8 LENGTH: 12124 base pairs
9 TYPE: nucleic acid
10 STRANDEDNESS: single
11 TOPOLOGY: linear
12 MOLECULE TYPE: DNA (genomic)
13 ORIGINAL SOURCE:
14 ORGANISM: Cucumis sativus
15 INDIVIDUAL ISOLATE: Cucumber Chitinase Genomic DNA
16 IMMEDIATE SOURCE:
17 CLONE: pBSCucchrhts
18 US-08-444-803-36
19
20 Query Match 9.3%; Score 29.8; DB 2; Length 12124;
21 Best Local Similarity 49.7%; Pred. No. 10;
22 Matches 73; Conservative 1; Mismatches 73; Indels 0; Gaps 0;
23
24 QY 49 ACAACAAAAAGAAAGTTGTGACATGTGTAGGAGAGTGTACCCCTTACTCCCAT 108
25 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
26 Db 10697 ACAGCAAAAAAATAATATGCAAGGTTGAGTACCATGTGTCATGATGCTCATATATAT 10756
27 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
28 QY 109 CAAAAAATAATGATACATGCTTAAAGATTAAGGCATATATTTATCATATATGTTCTA 168
29 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
30 Db 10757 ATATATATATATATATATATTCCTCATATATAAAAAATGCAAAATCTTTAAATAATATACA 10816
31 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
32 QY 169 AAAGAGAAGAGAGAGAAAAATACTACTT 195
33 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
34 Db 10817 GAGATGAATTAATTAATAATAAGTATT 10843
35
36 RESULT 15
37 US-08-449-043-36
38 Sequence 36, Application US/08449043
39 Patent No. 5689044
40 GENERAL INFORMATION:
41 APPLICANT: Ryals, John A.
42 APPLICANT: Alexander, Danny C.
43 APPLICANT: Beck, James J.
44 APPLICANT: Duesing, John H.
45 APPLICANT: Friedrich, Leslie B.
46 APPLICANT: Goodman, Robert M.
47 APPLICANT: Harms, Christian
48 APPLICANT: Melms, Jr., Frederick
49 APPLICANT: Montoya, Alice
50 APPLICANT: Moyer, Mary B.
51 APPLICANT: Neuhaus, Jean-Marc
52 APPLICANT: Payne, George B.
53 APPLICANT: Sperison, Christoph
54 APPLICANT: Stinson, Jeffrey R.
55 APPLICANT: Ukens, Scott J.
56 APPLICANT: Ward, Eric R.
57 APPLICANT: Williams, Shericca C.
58 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
59 NUMBER OF SEQUENCES: 106
60 CORRESPONDENCE ADDRESS:
61 ADDRESSEE: CIBA-GEIGY Corporation
62 STREET: 7 Skyline Drive
63 CITY: Hawthorne
64 STATE: New York
65 COUNTRY: USA
66 ZIP: 10532
67 COMPUTER READABLE FORM:
68 MEDIUM TYPE: Floppy disk
69 COMPUTER: IBM PC compatible
70 OPERATING SYSTEM: PC-DOS/MS-DOS

```

[illegible]

Query Match 9.3%; Score 29.8; DB 2; Length 12124;
Best Local Similarity 49.7%; Pred. No. 10;

This Page Blank (uspto)

OM of: US-09-030-606-224 to: PIR_60:* out_format : pfs
 Date: Sep 25, 1999 6:28 PM

About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:

-MODEL=framet.n2p.model -DEV=x1p
 -O=/cgr2.1/USPRO.spool/US09030606/runat_24091999_171617_29825/app.query.fasta.1
 -DE_PIR_60 -QMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELDP=6.000
 -DEEXT=7.000 -START=1 -MATRIX=blonsum62 -TRANS=humand0.cdt
 -LIST=45 -DOCALLIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
 -OUTFM=pfs -NORM=stat -USER=US09030606 -NCPU=6 -ICPU=3 -WAIT
 -THREADS=1

Search information block:

Query: US-09-030-606-224
 Query length: 320
 Database: PIR_60:*
 Database sequences: 122810
 Database length: 40065486
 Search time (sec): 182.540000

Sequence	Strd Orig	ZScore	EScore	len	Documentation
PIR2:F64221	+	71.00	152.46	0.6265	141 Documentation
PIR2:S50912	+	66.50	126.77	2.45	971 translation initiation factor IF-3
PIR2:S54426	+	66.00	128.63	2.83	663 DNA topoisomerase (ATP-hydrolyz
PIR2:S73365	+	65.00	135.74	3.01	201 translation initiation factor I
PIR2:A60944	+	64.00	131.54	5.07	255 cytochrome b - Leishmania mexic
PIR2:C71697	+	63.50	131.99	5.87	208 hypothetical protein RP396 - R
PIR2:I48134	+	63.50	127.24	5.91	379 ubiquinol--cytochrome-c reduct
PIR2:I48180	+	63.50	127.24	5.91	379 ubiquinol--cytochrome-c reduct
PIR2:A44217	+	63.00	108.83	6.10	389 genome polyprotein - bovine vi
PIR2:A45554	+	63.00	119.29	6.94	895 tholtry complex polypeptide Rm
PIR2:T00379	+	62.00	121.26	8.01	603 KIA0640 protein - human (fragm
PIR2:NM5H	+	62.00	114.44	9.39	1234 complement factor H precursor
PIR2:A70122	+	61.50	127.35	10.64	208 glucose inhibited division prot
PIR2:B39218	+	61.50	121.18	10.75	454 nicotinic acetylcholine recept
PIR2:TC3053	+	61.00	114.11	12.60	959 ribonucleoside diphosphate red
PIR2:S50434	+	61.00	112.42	12.63	1188 hypothetical protein YEL025c -
PIR2:S15656	+	60.50	124.13	14.34	233 zein, 19k - maize
PIR2:I49369	+	60.50	120.28	14.44	379 ubiquinol--cytochrome-c reduct
PIR2:HT0362	+	60.50	115.53	14.55	692 glycogen phosphorylase - Aquif
PIR2:HT0199	+	60.50	114.18	14.59	821 outer membrane protein homolo
PIR2:A55138	+	60.50	108.51	14.73	1681 sodium channel mna2.3, volta
PIR2:S16671	+	59.50	121.83	16.68	269 Dd31 protein - slime mold (Dich
PIR2:F30010	+	59.50	116.73	19.48	444 NADH dehydrogenase (ubiquinone)
PIR2:DT0397	+	59.50	116.71	19.48	444 cation efflux system (ccx8-like)
PIR2:C49596	+	59.00	114.53	19.55	585 genome polyprotein - border dls
PIR2:A53342	+	59.00	113.48	22.65	517 antiophosphoribosyltransferase
PIR2:EA4328	+	59.00	113.48	22.69	577 glycine-tRNA ligase (EC 6.1.1.
PIR2:D12053	+	59.00	112.75	22.71	633 paraspore crystal protein B2 -
PIR2:S59428	+	59.00	111.93	22.75	702 probable membrane protein YDR22
PIR2:S15655	+	58.50	119.42	26.01	235 zein, 19k - maize
PIR2:S07595	+	58.50	118.74	26.04	256 hypothetical protein, 29.4k - C
PIR2:S35572	+	58.50	115.64	26.18	379 ubiquinol--cytochrome-c reduct
PIR2:I48138	+	58.50	115.64	26.18	379 ubiquinol--cytochrome-c reduct
PIR2:S50592	+	58.00	114.05	26.25	464 hypothetical protein YER089c -
PIR2:ET1816	+	58.00	117.84	30.20	248 probable outer membrane protein
PIR2:BA5714	+	58.00	115.07	30.35	352 trans-activating transcription
PIR2:SE3442	+	58.00	114.57	30.37	375 probable protein kinase YOL126c
PIR2:BD4015	+	58.00	114.08	30.40	399 hypothetical protein H10874 - H
PIR2:DS9869	+	58.00	113.22	30.44	445 spore cortex membrane protein
PIR2:T00317	+	58.00	104.75	30.88	1300 serine proteinase esp, extrac
PIR2:S26026	+	57.50	120.97	34.80	144 NADH dehydrogenase (ubiquinone)

PIR2:D64369 + 57.50 118.99 34.91 185 | hypothetical protein MJ0556
 PIR2:S06214 - 57.50 114.84 35.16 313 | NADH dehydrogenase (ubiquino
 PIR2:A40455 + 57.50 114.49 35.18 327 | complement factor H-related

seq_name: PIR2:F64221

seq_documentation_block:

translation initiation factor IF-3 - Mycoplasma genitalium (SGC3)
 C:Species: Mycoplasma genitalium
 C:Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 21-Aug-1998
 C:Accession: F64221
 R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.
 M.; Fuhmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,
 C.A.; Venter, J.C.
 Science 270, 397-403, 1995
 A:Title: The minimal gene complement of Mycoplasma genitalium.
 A:Reference number: A64200; MID:96026346
 A:Accession: F64221
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-141 <TIGR>
 A:Cross-references: GB:U39697; GB:L43967; NID:91045878; PID:91045881; TIGR:MG196
 A:Experimental source: strain G-37
 C:Genetics:
 A:Genetic code: SGC3
 C:Superfamily: translation initiation factor IF-3

alignment_scores:

Quality: 71.00 Length: 76
 Ratio: 1.821 Gaps: 3
 Percent Similarity: 51.316 Percent Identity: 31.579

alignment_block:

US-09-030-606-224 x F64221

Align seg 1/1 to: F64221 from: 1 to: 141

56 AAGAGAAAAGTTTGACATTGTAGTACGAGAGTGTACCCCTTACTGCC 105
 :::::||||| :::::||||| ::::: |||
 9 GtngtlysglnleuaspLeuVallelaIAPoanPrtHlyspr 25
 106 CATCAAAAAGATGATACATGTTAAGATGATRAAGGCAATATTT 155
 ||||| ||| ||| |||
 25 otleallys.....LeuLeuaspPheGlyArgTyrTrt 37
 37 yraspleu.....LysArglyLysArg..... 44
 156 ATCATATGTTTAAGAGAGAGAGAGAAATACTACTTCTCAATG 205
 || :::: ||||| ::::: |||
 206 GAAGCCCTTAAGGTCTTGTATGATGAGACACAAATGCGCGTCCA 255
 ::||| ||||| ::::: |||
 45 GlnAlaLysLysAsnGlnThrIleleGlnThrLysGluValVally 61
 256 TCCGCTTARAGTGCATGACTTGAC 283
 ||| ::::: ||||| ::::: |||
 61 sprThrIleAlaLysShisAspleuGln 70
 seq_name: PIR2:S50912
 seq_documentation_block:
 probable membrane protein YDR166c - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein YDR358.20c; hypothetical protein YD9489.01c
 C:Species: Saccharomyces cerevisiae
 C:Date: 10-Feb-1995 #sequence_revision 23-Aug-1996 #text_change 06-Feb-1998
 C:Accession: S50912; S57990
 R:Oliver, K.; Harris, D.
 submitted to the EMBL Data Library, January 1994
 A:Reference number: S50912
 A:Accession: S50912
 A:Molecule type: DNA
 A:Residues: 1-412
 A:Cross-references: EMBL:Z47813; NID:g642274; PID:g642275; MIPS:YDR166c

R.Murphy, L.; Richards, C.; Harris, D.
submitted to the EMBL Data Library, July 1995

A:Reference number: S57971

A:Accession: S57990

A:Molecule type: DNA

A:Residues: 81-971 <EMBL>

A:Cross-references: EMBL:250046; NID:g899393; PID:g899413; MIPS:YDR166C

A:Experimental source: strain AB972

C:Genetics:

A:Gene: SGD:SECS

A:Cross-references: SGD:S0002573; MIPS:YDR166C

A:Map position: 4R

C:Keywords: transmembrane protein

F:845-861/Domain: transmembrane #status predicted <TM>

alignment_scores:

Quality:	66.50	Length:	75
Ratio:	1.415	Gaps:	3
Percent Similarity:	62.667	Percent Identity:	33.333

alignment_block:

US-09-030-606-224/rev x S50912 ..

Align seg 1/1 to: S50912 from: 1 to: 971

```

286 CGTGTCCAGTCATGCACTTAAGAGAGAGGACGCCACATTGTGT 237
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
204 ArgThrThrThrPhelyleuLeuProleuMetaspasnrygluysil 220
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
236 CCTTCAGTATCAAGACCTTTAAGGGCTTCATTTTGAGAAAGTACTAT 187
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
220 eteuastrygluAlaThrlysrlyspheilegluleuasnllyserlyr 237
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
186 TTTCTCTCTCTCTCTTTAGAACATATGATAAATTTCCCTTATC 137
    |||:::|||||:::|||||:::|||||:::|||||:::|||||
237 heasnneupolysSerleu.....LysArgCys.LeuThrAs 249
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
136 CTTTAAACATGATTCATTTTGTGATGGAGGAGTAAAGGGTACAC 87
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
249 n...AanaSpheasngluPhelileilegluYrserlysglyneuhrl 265
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
86 TCCCTACTACATGTCACAAACT 64
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
265 euArgrArgrArgrPheasnGlnSer 272
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||

```

seq_name: p1r2:S54426

seq_documentation_block:

DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) chain B - *Staphylococcus aureus*
C:Species: *Staphylococcus aureus*
C:Date: 06-Sep-1996 #sequence_revision 13-Mar-1997 #text_change 08-Sep-1997
C:Accession: S54426
R:Ferrero, L.; Cameron, B.; Manse, B.; Lagneaux, D.; Crouzet, J.; Famechon, A.; Blanche, M.; Microbiol. 13, 641-653, 1994
A:Title: Cloning and primary structure of *Staphylococcus aureus* DNA topoisomerase IV: a
A:Reference number: S54426; MUID:95089700
A:Accession: S54426
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-663 <FER>
A:Cross-references: EMBL:L25288; NID:g561878; PID:g561879
C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain B
C:Keywords: antibiotic resistance; ATP; DNA supercoiling; isomerase

alignment_scores:

Quality:	66.00	Length:	72
Ratio:	1.650 <td>Gaps:</td> <td>3</td>	Gaps:	3
Percent Similarity:	55.556 <td>Percent Identity:</td> <td>30.556</td>	Percent Identity:	30.556

alignment_block:

US-09-030-606-224 x S54426 ..

Align seg 1/1 to: S54426 from: 1 to: 663

```

43 ATAGAACACAAAAAAAGTTGTGACATGTGTAGGAGGTGTCT 92
    |||:::|||||:::|||||:::|||||:::|||||:::|||||
29 IleGlySerThrAspLysArgGlyLeuHisLeuValTyrGluLeu 45
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
93 ACCCTTACTCCCATCAAAAAAATGATGATCATGTGTAAGATATA 142
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
45 LaspaAsnSerValAspGluValLeuAsnGlyTyr..... 56
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
143 AGGCAATATTTATCATATGTTCTTAAAGAGAGAGAGAAATACTA 192
    |||:::|||||:::|||||:::|||||:::|||||:::|||||
57 ..GlyAsnGluLeuAspValThrile...AsnLysAspGlySerIleSer 71
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
193 CTTTCTCRAAATGGAGCCCTTAAAGTCTTGTGATCTGAGAGACAA 242
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
72 IleGluAspAsnGly.....ArgGlyMetProThrGlyThrHisly 85
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
243 ATGTGCCCGTCATCC 258
    |||:::|||||:::|||||:::|||||:::|||||:::|||||
85 sSerGlyLysProThr 90
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||

```

seq_name: p1r2:S73365

seq_documentation_block:

translation initiation factor IF-3 - *Mycoplasma pneumoniae* (ATCC 29342) (SGC3)
N:Alternate names: hypothetical protein C09_orf201
C:Species: *Mycoplasma pneumoniae*
A:Variety: ATCC 29342
C:Date: 26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
R:Himmelfreid, R.; Hilbert, H.; Plagens, H.; Plickl, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*
A:Reference number: S73367; MUID:97105885
A:Accession: S73365
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-201 <HIM>
A:Cross-references: EMBL:AE000005; GB:U00089; NID:g1673684; PID:g1673688
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Gene: infc
A:Genetic code: SGC3
C:Superfamily: translation initiation factor IF-3

alignment_scores:

Quality:	65.00	Length:	76
Ratio:	1.667 <td>Gaps:</td> <td>3</td>	Gaps:	3
Percent Similarity:	51.316 <td>Percent Identity:</td> <td>30.263</td>	Percent Identity:	30.263

alignment_block:

US-09-030-606-224 x S73365 ..

Align seg 1/1 to: S73365 from: 1 to: 201

```

56 AAGAAAAAGTTGTGACATGTGTAGGAGGTGTACCCCTTACTCC 105
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
69 GluGluLysGlnLeuAspLeuValIleGlySerAsnProAlaLysPr 85
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
106 CACCAAAAAAATGATGATGATGATTAAGGATTAAGGGCAATTTT 155
    |||:::|||||:::|||||:::|||||:::|||||:::|||||
85 olleValLys.....LeuLeuAspPheGlyArgTyrThrT 97
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
156 ATCATATGTTCTTAAAGAGAGAGAGAGAAATATCTACTTTCRAAAG 205
    |||:::|||||:::|||||:::|||||:::|||||:::|||||
97 yraspleu.....LysArgLysLysArg..... 104
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
206 GAAGCCCTTAAGGTCGTTGATGATGAGAGACAAATGTGCCCTCA 255
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||
105 GlnSerLysLysAsnGlnThrIleIleGlnIleLysGluValValVally 121
    ::|||:::|||||:::|||||:::|||||:::|||||:::|||||

```



```
217 TTAAAGGCTTCATTTGAGAAAGTAGATTTCTCTCTCTTT 168
    ||| :||| ||| ||| ||| ||| ||| ||| ||| |||
178 Pheheleaphehis.....PheValleupropheileil 189
167 AGAACATATGATAAATATTCCTTTATCCATTACCATGATTCATT 118
    ||| :||| ||| ||| ||| ||| ||| ||| ||| |||
189 ethAlaMeVal.....MetIleHsL 197
117 TTTTGTGATGGAGTAAAGGGGTACACACT..... 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
197 euLeupheleuHsLgIuThrGlySerAsnAsnProSerGlyLeuAsnSer 213
85 .....CCCTACTAC.....AATGTCAC 69
214 AsperasprySilePropheHsProTyTrThrIleLysAspIleLe 230
68 AAACCTTTTCTTTGTTGTTCCATTGGTTGTAACGTACTATTTCCT 20
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
230 uGlyLeuMetPheMetLeuPheAlaLeuMetMetLeuLeuPheSer 246
seq_name: p1r2:I48180
```

```
seq_documentation_block:
ubiquinol-cytochrome c reductase (EC 1.10.2.2) cytochrome b - American spiny rat (Isoth
C:Species: mitochondrion Isothrix bistriatus (American spiny rat)
C>Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 24-Jul-1998
C:Accession: I48180
R:da Silva, M.N.; Patton, J.L.
Mol. Phylogenet. Evol. 2, 243-255, 1993
A:Title: Amazonian phylogeography: mtDNA sequence variation in arboreal echymnid rodents
A:Reference number: A49605; MID:94184505
A:Accession: I48180
A:Status: preliminary; translated from GB/EMBL/DDJ
A:Molecule type: DNA
A:Residues: 1-379 <RES>
A:Cross-references: GB:I23355; NID:995846; PID:995847
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SSCI
C:Superfamily: cytochrome b; cytochrome b homology; cytochrome b6 homology; plastocuinol
C:Keywords: Chromoprotein; electron transfer; heme; iron; mitochondrion; oxidative phosph
F:I1-339/Domain: cytochrome b6 homology <CB6>
F:I1-209/Domain: cytochrome b6 homology <CB6>
F:221-339/Domain: plastocuinol-plastocyanin reductase 17K protein homology <17K>
```

```
alignment_scores:
    quality: 63.50      length: 83
    ratio: 1.587      gaps: 4
Percent Similarity: 48.193      Percent Identity: 27.711
```

alignment_block:

US-09-030-606-224/rev x I48180 ..

Align seg 1/1 to: I48180 from: 1 to: 379

```
217 TTAAAGGCTTCATTTGAGAAAGTAGATTTCTCTCTCTTT 168
    ||| :||| ||| ||| ||| ||| ||| ||| ||| |||
178 Pheheleaphehis.....PheValleupropheileil 189
167 AGAACATATGATAAATATTCCTTTATCCATTACCATGATTCATT 118
    ||| :||| ||| ||| ||| ||| ||| ||| ||| |||
189 ethAlaMeVal.....MetIleHsL 197
117 TTTTGTGATGGAGTAAAGGGGTACACACT..... 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
197 euLeupheleuHsLgIuThrGlySerAsnAsnProSerGlyLeuAsnSer 213
85 .....CCCTACTAC.....AATGTCAC 69
214 AsperasprySilePropheHsProTyTrThrIleLysAspIleLe 230
```

```
64 .....TTTTCTTTGTTGTTCCATTGGTTGTAACGTACTATTTCCT 20
    ||| :||| ||| ||| ||| ||| ||| ||| ||| |||
230 uGlyLeuMetPheMetLeuPheAlaLeuMetMetLeuLeuPheSer 246
seq_name: p1r1:A44217
```

seq_documentation_block:

genome polypotein - bovine viral diarrhea virus (strain SD-1)

C:Species: bovine viral diarrhea virus, BVDV

C>Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 15-May-1998

C:Accession: A44217

R: Deng, R.; Brock, K.V.

Virology 191, 867-879, 1992

A:Title: Molecular cloning and nucleotide sequence of a pestivirus genome, noncytopat

A:Reference number: A44217

A:Accession: A44217

A:Molecule type: genomic RNA

A:Residues: 1-3898 <DEN>

A:Cross-references: GB:M96751; NID:9289507; PID:9289508

A:Note: this polypotein may be cleaved into several mature proteins, including p20 p

and p75 protein; the cleavage sites are not reported

C:Superfamily: pestivirus genome polypotein; DEAD/H box helicase homology

C:Keywords: ATP; glycoprotein; P-loop; polypotein; RNA binding; zinc finger

F:2-234/Product: viral proteinase p20 #status predicted <VPT>

F:253-265/Region: hydrophobic

F:546-1115/Product: major envelope glycoprotein gp55 #status predicted <EGP>

F:547-565/Region: hydrophobic

F:556-670/Region: hydrophobic

F:675-694/Region: hydrophobic

F:1031-1046/Region: hydrophobic

F:1074-1099/Region: hydrophobic

F:1149-1164/Region: hydrophobic

F:1217-1238/Region: hydrophobic

F:1252-1269/Region: hydrophobic

F:1271-1292/Region: hydrophobic

F:1293-1304/Region: hydrophobic

F:1357-1373/Region: hydrophobic

F:1484-1512/Region: zinc finger

F:1815-2098/Domain: DEAD/H box helicase homology <DEAD>

F:1906-1911/Region: nucleotide-binding motif A (P-loop)

F:1910-1913/Region: DEAD motif B

F:2562-2582/Region: hydrophobic

F:272,281,296,335,365,370,413,487,597,809,878,922,990,1357,1419,1713,2134,2217,2494,2

```
alignment_scores:
    quality: 63.50      length: 125
    ratio: 0.992      gaps: 7
Percent Similarity: 51.200      Percent Identity: 24.000
```

alignment_block:

US-09-030-606-224/rev x A44217 ..

Align seg 1/1 to: A44217 from: 1 to: 3898

```
320 GTCCATGCGAGCTTAAAGTGAACACTTACCGTCCATGATGTC 271
    ||| :||| ||| ||| ||| ||| ||| ||| ||| |||
958 ValIleAlaThrAspIhrLysLeuGlyProMetProCysLysProHsG1 974
270 AACCTTAAGGAGATGGA.....CGGCACATTTGTGCTTCAG 230
    ||| :||| ||| ||| ||| ||| ||| ||| ||| |||
974 uIleIleSerSerGluGlyProIleGlyLysThrAlaCysThrPheAsn 990
229 TATCAAGACACCTTAAAGGCTCCATTTGAGAAAGTAGATTTCTCT 180
    ||| :||| ||| ||| ||| ||| ||| ||| ||| |||
991 TyrThrArgThrLeuLysAsnLysTyRPhGlu..... 1001
179 TCCCTCT.....CTTTAGACATATGATATAA.....TATGGC 145
    ||| :||| ||| ||| ||| ||| ||| ||| ||| |||
1002 ProAlaGspSerTyRPhGlnGlnTyRMetLeuLysGlyAspTyR.... 1016
144 CTCTATATCTTTAAACATGATCATATTTTGTGATGGAGGATAGG 95
```

```

1017 .....GlnTYRTPheAspLeuGluValThr 1025
94 GTACACACTCCCTACTACATGTCACAACTTTTC..... 59
1026 ASPHISHARGAspTYRPhelAGuSerIleLeuValValValAl 1042
58 .....TTTTGTGTCCTATTGGTGTGAAGTACTACTAT 25
1042 aLeuLeuGlyGlyArgTYRValLeuTPheLeuValThrMetVal 1059
24 TTCTTACACAGAGCGCTTCAGCG 2
1059 euSerGluGlnIlysalaserGly 1066

```

seq_name: p1r2:A45554

```

seq_documentation_block:
  rhoptry complex polypeptide RhopH3 - Plasmodium falciparum
  C:Species: Plasmodium falciparum
  C:Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 24-Feb-1995
  C:Accession: A45554
  R:Brown, H.J.; Coppel, R.L.
  Moll. Biochem. Parasitol. 49, 99-110, 1991
  A:Title: Primary structure of a Plasmodium falciparum rhoptry antigen.
  A:Reference number: A45554; MOID:92131054
  A:Contents: rhoptry
  A:Status: preliminary
  A:Molecule type: DNA
  A:Residues: 1-895 <BRO>
  A>Note: sequence extracted from NCBI backbone (NCBIN:77766, NCBI:77768)

```

```

alignment_scores:
  Quality: 63.00 Length: 84
  Ratio: 1.400 Gaps: 3
  Percent Similarity: 53.571 Percent Identity: 26.190

```

alignment_block:
US-09-030-606-224/rev x A45554 ..

Align seg 1/1 to: A45554 from: 1 to: 895

```

247 CACATTTGTCCTCAGTATCAACAGCCTTAGGCTCCATTGCA 198
|||||.....:|||||
533 HISValLeuValValIlySPHISrYrThrPhelSPHISrYrThG 549
197 GAAGTACTATTCTCTCTCTCT.....TAGAAC 163
|||||:|||||
549 uGluIleYrIlePhelLeuLysHISLeuIleIleLeuGlyLeuSera 566
162 ATATGATATAATATGCCCCCTTATCCCTTAACCATGATCCATTGTT 113
|||||:|||||
566 snLeuValSerSerAsnProSer.....SerProSer 576
112 TTGATGGGAGTAAAGGATACACCTCCTACTACATGTC..ACAA 66
|||||.....:|||||
577 PhesPThrIleIleGluPhelLeuValThrYrAsnLysGlySerG 593
65 CTTTTCCTTTGTTGTCCTATTGTTGTAAGTACTACTATTGTAACA 16
|||||:|||||
593 uLysPhelValLeuYrPhelIleSerIleIleSerValLeuYrIleAsn 610
15 AG 14
610 lu 610

```

seq_name: p1r2:T00379

```

seq_documentation_block:
  KIA00640 protein - human
  C:Species: Homo sapiens (man)

```

```

C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 01-Feb-1999
C:Accession: T00379
R:Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura,
DNA Res. 5, 169-176, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. X. The compl
A:Reference number: Z14142
A:Accession: T00379
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-603 <ISH>
A:Cross-references: EMBL:AB014540; NID:d1204295; PID:d1032576
C:Genetics:
A:Gene: KIA00640

```

```

alignment_scores:
  Quality: 62.50 Length: 63
  Ratio: 2.155 Gaps: 3
  Percent Similarity: 46.032 Percent Identity: 34.921

```

alignment_block:
US-09-030-606-224 x T00379 ..

Align seg 1/1 to: T00379 from: 1 to: 603

```

115 AAAATGATCATGCTTAAGATARAAGGC..... 147
|||||.....:|||||
231 LysGlnGlyrIlyrMetMetLys.....LysGlyHisArgArgLysAsnTr 245
148 .....AATTTTATCATATGTT.... 165
245 pThrGluArgTrpPhelValLeuLysProAsnIleIleSerYrYrValS 262
166 .....CTAAAGAGAGAGAGAGAAATATCTACTTTCTCRAAATGGA 207
|||||:|||||
262 ergLAspLeuLysAspLysLysGlyAspIleLeuLysGluAsnGlyS 278
208 AGCCCTTAAGCTGCTTGTGATGATGAGAGACACAAATGT 246
279 CysValGluSerLeuProAspLysAspLysLysLysCys 291

```

seq_name: p1r1:NBM5H

seq_documentation_block:

```

  Complement factor H precursor - mouse
  N:Alternate names: protein beta-1-H
  C:Species: Mus musculus (house mouse)
  C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 20-Mar-1998
  C:Accession: A26154; I49711; I49728
  R:Kristensen, T.; Tack, B.F.
  Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986
  A:Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in lengt
  A:Reference number: A26154; MOID:86233353
  A:Accession: A26154
  A:Molecule type: mRNA
  A:Residues: 1-1234 <KR1>
  A:Cross-references: GB:M12660; NID:g193724; PID:g387181
  R:Natsume-Sakai, S.; Nonaka, M.; Nonaka, M.; Harada, Y.
  J. Immunol. 144, 358-362, 1990
  A:Title: Demonstration of an unusual allelic variation of mouse factor H by the compl
  A:Reference number: I49711; MOID:90111033
  A:Accession: I49711
  A:Status: preliminary; translated from GB/EMBL/DBJ
  A:Molecule type: mRNA
  A:Residues: 1-18 <RES>
  A:Cross-references: GB:M31979; NID:g193726; PID:g193729
  R:Minoz-Canoes, P.; Tack, B.F.; Vlk, D.P.
  Biochemistry 28, 9891-9897, 1989
  A:Title: Analysis of complement factor H mRNA expression: Dexamethasone and IFN-gamma
  A:Reference number: I49728; MOID:90148935
  A:Accession: I49728
  A:Status: preliminary; translated from GB/EMBL/DBJ
  A:Molecule type: mRNA

```

A:Residues: 1-19 <RE2>

A:Cross-references: GB:J02891; NID:g193805; PID:g553926

C:Comment: Two codominant alleles of factor H are present in mice.

C:Comment: Factor H functions as a cofactor in the inactivation of C3b by serine proteinase (C5 convertase) in the alternative complement pathway.

C:Genetics:

A:Map position: 1

C:Superfamily: complement factor H; complement factor H repeat homology

C:Keywords: complement alternate pathway; duplication; glycoprotein; plasma

F:1-18/Domain: signal sequence status predicted <Sig>

F:19-123/Domain: complement factor H repeat homology <FH01>

F:21-80/Domain: complement factor H repeat homology <FH02>

F:85-141/Domain: complement factor H repeat homology <FH03>

F:146-205/Domain: complement factor H repeat homology <FH04>

F:210-267/Domain: complement factor H repeat homology <FH05>

F:246-248/Region: cell attachment (R-G-D) motif

F:267-320/Domain: complement factor H repeat homology <FH06>

F:325-385/Domain: complement factor H repeat homology <FH07>

F:389-442/Domain: complement factor H repeat homology <FH08>

F:448-505/Domain: complement factor H repeat homology <FH09>

F:509-564/Domain: complement factor H repeat homology <FH10>

F:569-623/Domain: complement factor H repeat homology <FH11>

F:629-683/Domain: complement factor H repeat homology <FH12>

F:690-743/Domain: complement factor H repeat homology <FH13>

F:752-802/Domain: complement factor H repeat homology <FH14>

F:808-861/Domain: complement factor H repeat homology <FH15>

F:867-931/Domain: complement factor H repeat homology <FH16>

F:936-989/Domain: complement factor H repeat homology <FH17>

F:994-1048/Domain: complement factor H repeat homology <FH18>

F:1053-1107/Domain: complement factor H repeat homology <FH19>

F:1114-1168/Domain: complement factor H repeat homology <FH20>

F:1212-1233/Domain: complement factor H repeat homology <FH21>

F:21-66,52-80,85-129,114-141,146-197,118-205,210-251,237-267,267-309,294-320,325-374,357-402,808-860,836-861,867-920,906-931,936-978,964-989,994-1037,1023-1048,1053-1096,1082-1082,808-860,836-861,867-920,906-931,936-978,964-989,994-1037,1023-1048,1053-1096,1082-1082,721,773,801,1030,1061,1223/Binding site: carbohydrate (Asn) (covalent) #status pre

alignment_scores:

Quality: 62.00 Length: 87

Ratio: 1.409 Gaps: 8

Percent Similarity: 50.575 Percent Identity: 29.885

alignment_block:

US-09-030-606-224 x NBMSH

Align seg 1/1 to: NBMSH from: 1 to: 1234

96 CCTTACTCCCA.....TCAAAAAAAAAATGAT 124

1057 ProHisValProAsnAlaThrIleValThrArgThrLysAsnLysTyrLeu 1073

125 ACATGGTTAAGATA.....RAAG 144

1073 uHisGlyAspArgValArgTyrGluCysAsnLysProLeuGluLeuPheG 1090

145 GCGAATATTTATCATATGTTTAAAGAGAGAGAGAAATACTACT 194

1090 LysGlnValGluValMetCys.....GluAsnGlyIle 1100

195 TTCTCRAAATGAAGCCCTTAAAGTCCTTATGATGAGACACAAAT 244

1101 TrpThrGlu...LysPro...LysCysArgAspSerThrGly...LysC 1114

245 GTGGCCCTCATCTCTTARAGTTCATGATGATGACACAGCAACTGT 294

1114 ysglyProPro. PropProIleAspAsnGlyAspIleThrSerLeuSerLe 1130

295 TGCAGTT 301

1130 uProVal 1132

seq_name: p1r2:A70122

seq_documentation_block:

glucose inhibited division protein B (gidb) homolog - Lyme disease spirochete

C:Species: Borrelia burgdorferi (Lyme disease spirochete)

C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998

C:Accession: A70122

R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh

son, D.; Peterson, J.; Kervatage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu

; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 380, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.

A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.

A:Reference number: A70100; MUID:98065943

A:Accession: A70122

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-208 <KLE>

A:Cross-references: GB:AE001128; GB:AE000783; NID:g2688057; PID:g2688060; TIGR:BB0177

A:Experimental source: strain B31

alignment_scores:

Quality: 61.50 Length: 91

Ratio: 1.255 Gaps: 4

Percent Similarity: 53.846 Percent Identity: 25.275

alignment_block:

US-09-030-606-224 x A70122

Align seg 1/1 to: A70122 from: 1 to: 208

47 GAACACAAACAAAGAAAGTTGTGACATGTAGTAGGAGGTGTACCC 96

68 GULLEASNRSGERGLUVALLEASPVAGLSERGLYALALYPHEPT 84

97 C.....GATARAAGGCAATATTTATCATATGTTTAAAGAGAG 177

84 OGlyIleIleLeuAlaIlePheAspSerSerArgLysTyrTyrLeuG 101

105 CCATCAACAAACAAAGCAATGCTTAAG..... 136

101 LuArgSerLysLysLysSerThrPheLeuLysMetIleLysLeuGlu 117

137GATARAAGGCAATATTTATCATATGTTTAAAGAGAG 177

118 AspleuGluAsnValLysIleLeuGluTyrGluIle.....GluYsgl 132

132 uLysLysLysTyrGluPheIleThrIleArgAlaPheArgAsnMetAsn 149

221GCTTGATACGTCAAGAC 238

149 LuTyrAlaLeuIleLeuLysAsn 156

seq_name: p1r2:B39218

seq_documentation_block:

nicotinic acetylcholine receptor alpha-5 chain precursor - chicken

C:Species: Gallus gallus (chicken)

C:Date: 30-Aug-1991 #sequence_revision 30-Aug-1991 #text_change 17-Mar-1999

C:Accession: B39218; S68498

R:Conturier, S.; Erkman, L.; Valera, S.; Runger, D.; Bertrand, S.; Boulter, J.; Ball

J. Biol. Chem. 265, 17560-17567, 1990

A:Title: alpha5, alpha3, and non-alpha3. Three clustered avian genes encoding neurona

A:Reference number: A39218; MUID:91009210

A:Accession: B39218

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-454 <CON>

A:Cross-references: GB:J05643

R:Ramirez-Latorre, J.; Yu, C.R.; Qu, X.; Perin, F.; Karlin, A.; Role, L.

Nature 380, 347-351, 1996

A:Title: Functional contributions of alpha-5 subunit to neuronal acetylcholine recept

A:Reference number: S68498; MUID:96176246

A;Accession: S68498

A;Molecule type: mRNA

A;ResIdues: 264-288 <RAM>

C; Superfamily: acetylcholine receptor

C; Keywords: glycoprotein; ion channel; neurotransmitter receptor; transmembrane protein

F:237-260/Domain: transmembrane #status predicted <TM1>

F:	268-286/Domain:
I:	269-270/Domain:
L:	271-272/Domain:
S:	273-274/Domain:
T:	275-276/Domain:
V:	277-278/Domain:

F:302-323/Domain: transmembrane #status predicted <TM3>
E:200-200/Domain: transmembrane #status predicted <TM2>

E:416-434/DomAin:	transmembrane	status	predicted	<TM>
F:302-323/DomAin:	transmembrane	status	predicted	<TM>

E:156-170 220-221 /misc/bonds: #status predicted <TM4>
F:416-434/Domain: transmembrane #status predicted <TM4>

F;169/Blinding site: carbohydrate (Asn) (covalent) #status predicted

alignment_scores:

Quality:	61.50	Length:	102
----------	-------	---------	-----

Ratio:	1.183	Gaps:	4
--------	-------	-------	---

Similarity:	50.980	Percent Identity:	29.412
-------------	--------	-------------------	--------

alignment_block:

506-224/rev x B39218

Align seg 1/1 to: B39218 from: 1 to: 454

[illegible]

seq_name: p1r2:T03053

seq_documentation_block:

ribonucleoside-diphosphate reductase large chain homolog - Chilo iridescent virus seq_documentation_block:

C: Spectra: Chl *a* fluorescence

C/Species: chilo iridescent virus
C/Date: 24-Mar-1998 #sequence reversion 24-Mar-1998

C;Date: 24-Mar-1999

C;Accession: T03053

R; Bahr, U.; Tldona, C.A.; Dar

Virus Genes 15, 235-245, 1997

A;Title: The DNA sequence

A:Reference number:

A: Accession: T03053
A: Reference number: Z14834

A;Accession: T03053
A>Status: Preliminary

A;Status: preliminary;

A;Molecule type: DNA

Percent Similarity: 48.000 Percent Identity: 30.667

alignment_block:

US-09-030-606-224/rev x T03053

Align seg 1/1 to: T03053 from: 1 to: 959

```

245  CATTTGTGTCCTTCAGATCAAAAGCAGCTTTAAAGGCTTCATTTGAGAA 196
372  .....LysL 373
195  AACTAGATTTTCTCTCTCTCTCTTA.....GAAATATATATA 155
373  eullleuysPheasPleuProthleuPheglYuaSerglunhlaSple 389
154  AAATATTTGCCCTTTATCTCTTAACCAAGTATTCATTTTCTTTGATGG 105
390  LysTyr.....ProtyrlhnhlaSgryPhePheCysgl 400
104  GAGATTAAGGGGACACACTCCCTAC 80
400  yaap...glYhrylYhrylLysTyr 407

```


SO SEQUENCE 663 AA; 74364 MW; E3682180 CRC32;

alignment_scores:

Quality: 66.00 Length: 72
Ratio: 1.650 Gaps: 3
Percent Similarity: 55.556 Percent Identity: 30.556

alignment_block:

US-09-030-606-224 x PARE_STAAU ..

Align seg 1/1 to: PARE_STAAU from: 1 to: 663

```

43 ATAGACACAAAGAAAGTTGTCATGTAGTACGAGGTGT 92
|||||:|||||:|||||:|||||:|||||:|||||:
29 IleglySerThrAspLysArgGlyLeuHisHisLeuValTyrGluIleVa 45
93 ACCCTTACTCCCATCAAAAAAATGATATCATGTTAAGGATARA 142
|||:|||||:|||||:|||||:|||||:|||||:
45 LaspAsnSerValAspGluValIleuAsnGlyTyr..... 56
143 AGCGAATATTTATCATATGTTCTTAAAGAGAGAGAAATACTA 192
|||||:|||||:|||||:|||||:|||||:|||||:
57 ..GlyAsnGluIleAspValThrIle...AsnLysAspGlySerIleSer 71
193 CTTTCTCRAAATGAAAGCCCTTAAGTGCTGTGATGAGAGACACA 242
|||||:|||||:|||||:|||||:|||||:|||||:
72 IlegluAspAsnGly.....ArgGlyMetProThrGlyIleHisIly 85
243 ATGTGGCCGTCATCC 258
|:|||||:|||||:|||||:|||||:
85 SserGlyLysProThr 90

```

seq_name: SwissProt_37:IF3_MYCPN

seq_documentation_block:

```

ID IF3_MYCPN STANDARD; PRT; 201 AA.
AC P78024;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE TRANSLATION INITIATION FACTOR IF-3.
GN IFN3.
OS MYCOPLASMA PNEUMONIAE.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
OC MYCOPLASMACTACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE: 97105885.
RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
RA HERRMANN R.
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RT NUCLEIC ACIDS RES. 24:4420-4449(1996).
CC -1- FUNCTION: IF-3 BINDS TO THE 30S RIBOSOMAL SUBUNIT AND SHIFTS THE
CC EQUILIBRIUM BETWEEN 70S RIBOSOMES AND THEIR 50S AND 30S SUBUNITS IN
CC FAVOR OF THE FREE SUBUNITS, THUS ENHANCING THE AVAILABILITY OF 30S
CC SUBUNITS ON WHICH PROTEIN SYNTHESIS INITIATION BEGINS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE IF-3 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000005; GI673688; -.
CC PROSITE: PS00938; IF3; 1.

```

DR PFAM: PF00707; IF3; 1.

DR HSP: P03000; ITF.

KW INITIATION FACTOR; PROTEIN BIOSYNTHESIS.

SO SEQUENCE 201 AA; 23145 MW; 1FC1283C CRC32;

alignment_scores:

Quality: 65.00 Length: 76
Ratio: 1.667 Gaps: 3
Percent Similarity: 51.316 Percent Identity: 30.263

alignment_block:

US-09-030-606-224 x IF3_MYCPN ..

Align seg 1/1 to: IF3_MYCPN from: 1 to: 201

```

56 AAGAAAAAGTTGTGACATTTAGTAGGAGGTGTACCCCTTACTCC 105
|||||:|||||:|||||:|||||:|||||:|||||:
69 GluGluLysGlnLeuAspLeuValIleuIleGlySerAsnProAlaLysPr 85
106 CATCAAAAAAATGATACATGTTAAGGATARAAGGCATATTTT 155
|||||:|||||:|||||:|||||:|||||:|||||:
85 OileValLys.....LeuLeuAspPheGlyArgTyrThr 97
156 ATCATATGTTCTTAAAGAGAGAGAAATACTACTTCTCRAAATG 205
||:|||||:|||||:|||||:|||||:
97 YrasPleu.....LysArgLysLysArg..... 104
206 GAAGCCCTTAAGGTGCTTGTATCTGACAGACACAATGTGCGCTCA 255
|||||:|||||:|||||:|||||:|||||:|||||:
105 GlnSerLysLysAsnGlnThrIleIleGlnIleLysGlnValValVally 121
256 TCCCTCTTARAGTTGCATGACTTGAC 283
|||:|||||:|||||:|||||:|||||:
121 sProThrIleAlaLysHisAspLeuGln 130

```

seq_name: SwissProt_37:CYB_CHEMA

seq_documentation_block:

ID CYB_CHEMA STANDARD; PRT; 379 AA.

AC Q34169;

DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE CYTOCHROME B (EC 1.10.2.2).

GN MTCYB OR COB OR CYTB.

OS CHEIROGALEUS MAJOR (LEMUR).

OC MITOCHONDRION.

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC PRIMATES; STREPSIRHINI; CHEIROGALEIDAE; CHEIROGALEUS.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 96209867.

RA YODER A.D., CARTMILL M., RUVOLO M., SMITH K., VILGALYS R.;

RT "Ancient single origin for Malagasy primates."

CC PROC. NATL. ACD. SCI. U.S.A. 93:5122-5126(1996).

CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE

CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A

CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL

CC COUPLED TO ATP SYNTHESIS.

CC -1- CATALYTIC ACTIVITY: O₂(H⁺) + 2 FERRICYTOCHROME C = O +

CC 2 FERROCYTOCHROME C.

CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY

CC BOUND TO THE PROTEIN.

CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,

CC CYTOCHROME C1 AND THE RIESKE PROTEIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

```

CC      This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/1sb-sib.ch).
CC      or send an email to license@1sb-sib.ch.
CC      -----
DR      EMBL; M96751; G395508; .
DR      PIR; A44217; A44217.
DR      PROSITE; PS00531; RNASE_T2_2; UNKNOWN_1.
DR      PFM; PF00271; helicase_C; 1.
KW      POLYPEPTIDE; GLYCOPROTEIN; HELICASE; SERINE PROTEASE; HYDROLASE.
FT      CHAIN 1 2270
FT      CHAIN 2 21063
FT      CHAIN 3 3898
FT      CHAIN 4 755
FT      DOMAIN 690 755
FT      ACT_SITE 1658 1658
FT      ACT_SITE 1695 1695
FT      ACT_SITE 1752 1752
FT      CARBOHYD 272 272
FT      CARBOHYD 281 281
FT      CARBOHYD 296 296
FT      CARBOHYD 335 335
FT      CARBOHYD 365 365
FT      CARBOHYD 370 370
FT      CARBOHYD 413 413
FT      CARBOHYD 487 487
FT      CARBOHYD 597 597
FT      CARBOHYD 809 809
FT      CARBOHYD 878 878
FT      CARBOHYD 922 922
FT      CARBOHYD 990 990
FT      CARBOHYD 1357 1357
FT      CARBOHYD 1419 1419
FT      CARBOHYD 1451 1451
FT      CARBOHYD 1713 1713
FT      CARBOHYD 2134 2134
FT      CARBOHYD 2217 2217
FT      CARBOHYD 2494 2494
FT      CARBOHYD 2682 2682
FT      CARBOHYD 2751 2751
FT      CARBOHYD 2891 2891
FT      CARBOHYD 2988 2988
FT      CARBOHYD 3688 3688
FT      CARBOHYD 3777 3777
FT      CARBOHYD 3793 3793
FT      SEQUENCE 3898 AA; 437800 MW; A562145C CRC32;
CC      -----
alignment_scores:
      Quality: 63.50      Length: 125
      Ratio: 0.992      Gaps: 7
Percent Similarity: 51.200      Percent Identity: 24.000

alignment_block:
US-09-030-606-224/rev x POLG_BVDVS ..
Align seg 1/1 to: POLG_BVDVS from: 1 to: 3898

320 GTCACAAATCGKAGTYTAACGTCAACAGTACGTCGCAAGTCATGC 271
||||| : : : : : ||||| : : : : : |||
958 VelllealThrAspThrLysIleuGlYPrometProCysLysProHisGl 974
||||| : : : : : ||||| : : : : : |||
270 AACTYTAAGGAGGATGAG.....CGGCACATTGTGTGCTCTCAG 230
: : : : : : : : : : : : : : : : : : : : : : : : : : :
974 uilleleserIseuGlYProIleGlLysThrAlacysThrPheAsn 990
: : : : : : : : : : : : : : : : : : : : : : : : : : :
229 TATCAAAACACCTTAAAGGCTTCATTATYGAAGAAAGTATTTCTCT 180
||||| : : : : : ||||| : : : : : |||
991 TYRThrArThrLeuLysAsnLysTyrPheGlu..... 1001

```

```

179 TCCTCT.....CTTAGACATATGATAAA.....TATGCC 145
1002 .Protagaspserlyrphelglintyrmelueylsglyaspyr... 1016
144 CTTTATCTTTAACCATGATATTCATTTTGTGATGGGAGTAAAGG 95
1017 .....Gintyrrtrpheaaplenuvalthr 1025
94 GTACACATCCCTACTACATGACAAACTTTTC..... 59
1026 AsphlshsavgaspyrPhealaguserileleuvalvalval 1042
58 .....TTTTGTCTCTATGTTGTTGACTGACTAT 25
1042 aleuenuglylarytyrvalleutrpheuleuvalthrlymetval 1059
24 TTTCTAACAGACGCTTCAGG 2
1059 eusergluglntylsalasergly 1066

seq_name: SwissProt_37:ACH5_CHICK

seq_documentation_block:
ID ACH5_CHICK STANDARD; PRT; 454 AA.
AC P26152;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DE 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE NEURONAL ACETYLCHOLINE RECEPTOR PROTEIN, ALPHA-5 CHAIN PRECURSOR.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
NOGONATHE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LESHORN; TISSUE-BRAIN;
RX MEDLINE; 91009210.
RA COUTURIER S., EKKMAN L., VALERA S., RUNGER D., BERTRAND S.,
BOULTER J., BALLYET M., BERTRAND D.;
RT "Alpha 5, alpha 3, and non-alpha 3. Three clustered avian genes
encoding neuronal nicotinic acetylcholine receptor-related
subunits."
RT J. BIOL. CHEM. 265:17560-17567(1990).
CC -1- FUNCTION: AFTER BINDING ACETYLCHOLINE, THE ACHR RESPONDS BY AN
EXTENSIVE CHANGE IN CONFORMATION THAT AFFECTS ALL SUBUNITS AND
LEADS TO OPENING OF AN ION-CONDUCTING CHANNEL ACROSS THE PLASMA
MEMBRANE.
CC -1- SUBUNIT: NEURONAL ACHR SEEMS TO BE COMPOSED OF TWO DIFFERENT
TYPE OF SUBUNITS: ALPHA AND NON-ALPHA (ALSO CALLED BETA). A
FUNCTIONAL RECEPTOR SEEMS TO CONSIST OF TWO ALPHA-CHAINS AND
THREE NON-ALPHA CHAINS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL: J05642; G311047; -
DR PIR: B39218; B39218.
DR PROSITE: PS00236; NEUROTR_ION_CHANNEL; 1.
DR PFAM: PF00065; neu_chan; 2.
KM RECEPTOR: POSTSYNAPTIC MEMBRANE; IONIC CHANNEL; GLYCOPROTEIN; SIGNAL;
TRANSMEMBRANE; MULTIGENE FAMILY.
FT CHAIN 1 29 POTENTIAL.
FT SIGNAL 1 29
FT DOMAIN 30 454 ACETYLCHOLINE RECEPTOR PROTEIN, ALPHA-5.
FT TRANSMEM 236 260 EXTRACELLULAR.
FT TRANSMEM 268 285 POTENTIAL.
FT TRANSMEM 302 323 POTENTIAL.

```

```

FT DOMAIN 324 415 CYTOPLASMIC.
FT TRANSMEM 416 435 POTENTIAL.
FT DISULFID 156 170 BY SIMILARITY.
FT DISULFID 220 221 ASSOCIATED WITH RECEPTOR ACTIVATION
(BY SIMILARITY).
FT CARBOHYD 55 55 POTENTIAL.
FT CARBOHYD 169 169 POTENTIAL.
FT CARBOHYD 215 215 POTENTIAL.
SQ SEQUENCE 454 AA; 51957 MW; 4A4AD7C5 CRC32;

alignment_scores:
Quality: 62.50 Length: 102
Ratio: 1.202 Gaps: 4
Percent Similarity: 50.980 Percent Identity: 29.412

alignment_block:
US-09-030-606-224/rev x ACH5_CHICK ..
Align seg 1/1 to: ACH5_CHICK from: 1 to: 454

315 AATGCKAGATTAAACTGCAACAGTTACGTCGACATGCACTY 266
||||| : : : : : ||| : : : : :
201 Asnglyluptrguilevalthralthrlyserlysglyasnaryth 217
265 TAAGGAGATGAGCGCCACATTTGTGCTCGATGTC.....A 225
||| : : : : : ||| : : : : :
217 rapselgyscystpyrProphnevalthrySerheilelearya 234
||| : : : : : ||| : : : : :
224 AAGCAGCTTTAAGGCGCTTCATTTYGAGAAAGTAGTATTTCTCTCT 175
:: : : : : ||| : : : : :
234 rgleupProleuphe.....Tyrthleupheleu 243
||| : : : : : ||| : : : : :
174 CTCTTTTGAACATATGATAAATATTTGCCCTTTATCTTAAACATGT 125
: : : : : ||| : : : : :
244 lleleProcytle.....GlyleuSerPheleuthVa 255
||| : : : : : ||| : : : : :
124 ATCATTTTGTGATGGGAGTAAAGGCTACACACTCCCTACTACAA 75
||| : : : : : ||| : : : : :
255 lleuvalPheTyr.....LeuProSerang 264
||| : : : : : ||| : : : : :
74 TGTCACAACTTTTCTTTGTGTTGCTTATGTTGTTGTTGTTGTTGTT 25
||||| : : : : : ||| : : : : :
264 luallaglylserleucylserleuvalleuvalserleuthval 280
||| : : : : : ||| : : : : :
24 TTTCTA 19
|||||
281 Pheleu 282

seq_name: SwissProt_37:GIBB_BORBU

seq_documentation_block:
ID GIBB_BORBU STANDARD; PRT; 208 AA.
AC P53363; 051197;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLUCOSE INHIBITED DIVISION PROTEIN B.
GN GIBB OR BB0177.
OS BORRELIA BURGDORFERI (LYME DISEASE SPIROCHETE).
OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; BORRELIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-212;
RA OLD I.G.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 35210 / B31;
RX MEDLINE; 98065943.
RA FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,

```

RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
 RA VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
 RA UTTERBACK T., WATHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
 RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
 RA SMITH H.O., VENTER J.C.,
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia
 burgdorferi.";
 RL NATURE 390:580-586(1997).
 [3]
 RP SEQUENCE OF 1-8 FROM N.A.
 RC STRAIN-212;
 RX MEDLINE: 93146383;
 RA OLD I.G., MACDOUGALL J., SAINT-GIRONS I., DAVIDSON B.E.;
 RT "Mapping of genes on the linear chromosome of the bacterium Borrelia
 burgdorferi: possible locations for its origin of replication.";
 RL FEMS MICROBIOL. LETT. 78:245-250(1992).
 CC -1- FUNCTION: NOT KNOWN
 CC -1- SIMILARITY: BELONGS TO THE GIDB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X95668; E223226;
 DR EMBL: AJ003222; E1216365;
 DR EMBL: X96434; E228293;
 DR EMBL: AE001128; G268060;
 DR EMBL: Z12160; E223419;
 DR TIGR: B80177;
 DR CONFLICT 75 84 LDVSGAGHP -> PWTLEVLDFR (IN REF. 1).
 FT CONFLICT 93 93 S -> T (IN REF. 1).
 FT CONFLICT 146 146 N -> S (IN REF. 1).
 FT SEQUENCE 208 AA; 24177 MM; 2B33BC90 CRC32;

alignment_scores:
 Quality: 61.50 Length: 91
 Ratio: 1.255 Gaps: 4
 Percent Similarity: 53.846 Percent Identity: 25.275

alignment_block:
 US-09-030-606-224 x GIDB_BORBU
 Align seg 1/1 to: GIDB_BORBU from: 1 to: 208

47 GAACACAAAAAGAAAGTTGTGACATGTAGGAGAGTGTACCC 96
 ||| : : : : : ||| : : : : :
 68 GATLEANSRPSRSGIUALLEUASPVALIGYSEGLYALAGLYPHEPR 84
 97 C.....TTACTCC 104
 84 OGILYILELLEUALLEPHEASPSERSEARGLYTYRTREULENG 101
 105 CCATCAAAAAAATGATGATGTTAAAG..... 136
 101 LUARGSERLYSLYSETHPHEULYSMETLIELYSLLEUGLUEN 117
 137GATRAAGGCAATATTTATCATATGTTCTAAAGAGAG 177
 118 Aspleuglunsnvallyslleleuglutyrciulle.....GIUYSGL 132
 178 GAAGAGAAATACACTTCTCRAAATGGAAGCCCTTAAAGCT..... 220
 132 ULYSLYSLYSLYTCIGUPHEITHTIRILEARGALAPHEARGSNMELASNG 149
 221GCTTGATCTGAGAGAC 238
 149 LUTYRALALEULLEULYSASN 156

seq_name: SwissProt_37:CFAH_MOUSE
 seq_documentation_block:
 ID CFAH_MOUSE STANDARD; PRT; 1234 AA.
 AC P06909;
 DT 01-JAN-1988 (REL. 06, CREATED)
 DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
 DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE COMPLEMENT FACTOR H PRECURSOR (PROTEIN BETA-1-H).
 GN CPH.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 86233353;
 RA KRISTENSEN T., TACK B.F.;
 RT "Murine protein H is comprised of 20 repeating units, 61 amino acids
 RT in length.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 83:3963-3967(1986).
 RN [2]
 RP SEQUENCE OF 1-19 FROM N.A.
 RC STRAIN-BAB/C;
 RX MEDLINE: 90148935;
 RA MUNOZ-CANOVES P., TACK B.F., VIK D.P.;
 RT "Analysis of complement factor H mRNA expression: dexamethasone and
 RT IFN-gamma increase the level of H in L cells.";
 RL BIOCHEMISTRY 28:9891-9897(1989).
 RN [3]
 RP SEQUENCE OF 1-18 FROM N.A.
 RX MEDLINE: 90111033;
 RA NATSUME-SAKAI S., NONAKA M., HARADA Y.N., SHREFFLER D.C.,
 RA MORIMAKI K.;
 RT "Demonstration of an unusual allelic variation of mouse factor H by
 RT the complete cDNA sequence of the H.2 allotype.";
 RL J. IMMUNOL. 144:358-362(1990).
 CC -1- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF
 CC C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE
 CC C3BB COMPLEX (C3 CONVERTASE) AND THE (C3)NBB COMPLEX (C5
 CC CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.
 CC -1- POLYMORPHISM: TWO CODOMINANT ALLELES OF FACTOR H ARE PRESENT IN
 CC MICE.
 CC -1- SIMILARITY: CONTAINS 20 SUSHI (SCR) REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: M12660; G387181;
 DR EMBL: J02891; G553926;
 DR EMBL: M31979; G193729;
 DR PIR: A26154; NBMNH.
 DR MGD: MGI:88385; CPH.
 DR PFM: PRO0084; sush1; 20.
 DR HSSP: P08603; IHTL.
 KW COMPLEMENT ALTERNATE PATHWAY; PLASMA; GLYCOPROTEIN; REPEAT; SUSHI;
 KW SIGNAL.
 FT SIGNAL. 1 18 BY SIMILARITY.
 FT CHAIN 19 1234 COMPLEMENT FACTOR H.
 FT DOMAIN 20 1224 20 X SUSHI (SCR) REPEATS.
 FT REPEAT 20 81 SUSHI 1.
 FT REPEAT 84 142 SUSHI 2.
 FT REPEAT 145 206 SUSHI 3.
 FT REPEAT 209 263 SUSHI 4.
 FT REPEAT 266 321 SUSHI 5.
 FT REPEAT 324 386 SUSHI 6.
 FT REPEAT 388 443 SUSHI 7.
 FT REPEAT 447 506 SUSHI 8.
 FT REPEAT 508 565 SUSHI 9.

```

FT REPEAT 568 623 SUSHI 10.
FT REPEAT 628 684 SUSHI 11.
FT REPEAT 689 744 SUSHI 12.
FT REPEAT 751 803 SUSHI 13.
FT REPEAT 807 862 SUSHI 14.
FT REPEAT 866 932 SUSHI 15.
FT REPEAT 935 990 SUSHI 16.
FT REPEAT 993 1049 SUSHI 17.
FT REPEAT 1052 1108 SUSHI 18.
FT REPEAT 1113 1169 SUSHI 19.
FT REPEAT 1171 1234 SUSHI 20.
FT DISULFID 21 66 BY SIMILARITY.
FT DISULFID 52 80 BY SIMILARITY.
FT DISULFID 85 129 BY SIMILARITY.
FT DISULFID 114 141 BY SIMILARITY.
FT DISULFID 146 192 BY SIMILARITY.
FT DISULFID 178 203 BY SIMILARITY.
FT DISULFID 210 251 BY SIMILARITY.
FT DISULFID 237 262 BY SIMILARITY.
FT DISULFID 267 309 BY SIMILARITY.
FT DISULFID 294 320 BY SIMILARITY.
FT DISULFID 325 374 BY SIMILARITY.
FT DISULFID 357 385 BY SIMILARITY.
FT DISULFID 389 431 BY SIMILARITY.
FT DISULFID 416 442 BY SIMILARITY.
FT DISULFID 448 494 BY SIMILARITY.
FT DISULFID 477 505 BY SIMILARITY.
FT DISULFID 509 553 BY SIMILARITY.
FT DISULFID 536 564 BY SIMILARITY.
FT DISULFID 569 610 BY SIMILARITY.
FT DISULFID 597 622 BY SIMILARITY.
FT DISULFID 629 672 BY SIMILARITY.
FT DISULFID 658 683 BY SIMILARITY.
FT DISULFID 690 732 BY SIMILARITY.
FT DISULFID 718 743 BY SIMILARITY.
FT DISULFID 752 791 BY SIMILARITY.
FT DISULFID 780 802 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 861 BY SIMILARITY.
FT DISULFID 867 920 BY SIMILARITY.
FT DISULFID 906 931 BY SIMILARITY.
FT DISULFID 936 978 BY SIMILARITY.
FT DISULFID 964 989 BY SIMILARITY.
FT DISULFID 994 1037 BY SIMILARITY.
FT DISULFID 1023 1048 BY SIMILARITY.
FT DISULFID 1053 1096 BY SIMILARITY.
FT DISULFID 1082 1107 BY SIMILARITY.
FT DISULFID 1114 1157 BY SIMILARITY.
FT DISULFID 1143 1168 BY SIMILARITY.
FT DISULFID 1172 1223 BY SIMILARITY.
FT DISULFID 1206 1233 BY SIMILARITY.
FT CARBOHYD 676 676 POTENTIAL.
FT CARBOHYD 721 721 POTENTIAL.
FT CARBOHYD 721 721 POTENTIAL.
FT CARBOHYD 773 773 POTENTIAL.
FT CARBOHYD 801 801 POTENTIAL.
FT CARBOHYD 1030 1030 POTENTIAL.
FT CARBOHYD 1061 1061 POTENTIAL.
FT CARBOHYD 1225 1225 POTENTIAL.
SQ SEQUENCE 1234 AA; 139082 MW; F0F4213F CRC32;

```

alignment_scores:

```

Quality: 62.00 Length: 87
Ratio: 1.409 Gaps: 8
Percent Similarity: 50.575 Percent Identity: 29.885

```

alignment_block:

US-09-030-606-224 x CFAH_MOUSE

Align seg 1/1 to: CFAH_MOUSE from: 1 to: 1234

```

96 CCTACTGCGCA.....TCMAAAAAAAAAATGGAT 124
||||: |||

```

```

1057 ProHsValProAsnAlaThrIleValThrArgThrLysAsnLysTyrLe 1073
125 ACATGCTAAAGGATA.....RAAG 144
1073 uHsGlyAspArgValArgTyrGluCysAsnLysProLeuGluLeuPheG 1090
145 GCATATATTATCATATGCTCTAAAGAGAGAGAGAAATACTACT 194
1090 lYgIlnValGluValMetCys.....GluAsnGlyIle 1100
195 TTCCTCRAATGGAAGCCCTTAAGGTGCTTGATACATGAGACACAAT 244
1101 TrpHrglu...LysPro...LysCysArgAspSerThrIle...LysC 1114
245 GTGCGCCGTCATCTCTTACCTTACCTGATGACATGACACGCTACTGT 294
1114 ySgIlyProPro.ProPoiIleAspAsnGlyAspIleThrSerLeuSerLe 1130
295 TGCAGT 301
1130 uProVal 1132

```

seq_name: SwissProt_37.ZEAD_MAIZE

seq_documentation_block:

ID ZEAD_MAIZE STANDARD; PRT; 233 AA.

```

AC P244F0;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ZEIN-ALPHA PRECURSOR (19 KD) (PMS2).
GN ZEMS2.
OS ZEA MAIS (MAIZE).
OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA; POALES;
OC POACEAE; ZEA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRATIN-CV. A619;
RX MEDLINE; 90060774.
RA OUAVIDE T.J.A., BROWN J.W.S., FEIX G.;
RL "Analysis of distal flanking regions of maize 19-kDa zein genes.";
RL GENE 80:249-257(1989).
RN [2]
RP SEQUENCE FROM N.A.
RA LANGRIDGE P., BROWN J.W.S., PINTOR-TORO J.A., FEIX G.;
RL "Expression of zein genes in Acetabularia mediterranea.";
RL EUR. J. CELL BIOL. 39:257-264(1985).
CC -1- FUNCTION: ZEINS ARE MAJOR SEED STORAGE PROTEINS.
CC -1- THE ALPHA ZEINS OF 19 KD AND 22 KD ACCOUNT FOR 70% OF THE TOTAL
CC ZEIN FRACTION. THEY ARE ENCODED BY A LARGE MULTIGENE FAMILY.
CC -1- STRUCTURALLY, 22K AND 19K ZEINS ARE COMPOSED OF NINE ADJACENT,
CC TOPOLOGICALLY ANTIPARALLEL HELICES CLUSTERED WITHIN A DISTORTED
CC CYLINDER.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC
CC EMBL; X58700; G22448; -
CC EMBL; X59526; G16073; ALT_TERM.
CC PIR; S15656; S15656.
CC MAIZEDB; 58096; -
CC SEED STORAGE PROTEIN. REPEAT: MULTIGENE FAMILY; SIGNAL.
FT SIGNAL 1 21
FT CHAIN 22 233 ZEIN-ALPHA.
SQ SEQUENCE 233 AA; 25515 MW; 0292BE85 CRC32;

```

alignment_scores:
 Quality: 60.50 Length: 66
 Ratio: 1.681 Gaps: 1
 Percent Similarity: 54.545 Percent Identity: 30.303

alignment_block:

US-09-030-606-224/rev x ZEAD_MAIZE ..

Align seg 1/1 to: ZEAD_MAIZE from: 1 to: 233

```

209 CTCCTGATTTGAGAAAGTACTATTTCTCTCTCTCTTTAGACATA 160
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:
169 TLePProPheserGlnLeuAlaaspValSerProAlaIalaphLeuThrG1 185
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:
159 TGATTAATATTTGCGCTTATACCTTTAACCATGATATTCATTTT 110
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:
185 nGlnGlnLeuLeuProPhetYrLeuHsAlaMet..... 136
109 GATGGGAGTAAAGGGGTACACACTCCCTACTACATGTCACAACTTTT 60
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:
197 .....ProAsnAlaGlyThrLeuLeuGlnLeuGlnLeuLeu 209
59 CTTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 12
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:
210 ProPhesAsnGlnLeuAlaLeuThrAsnSerThrValPhetYrGlnGln 225

```

seq_name: SwissProt_37:YEC5_YEAST

seq_documentation_block:

ID YEC5_YEAST STANDARD; PRT; 1188 AA.

AC P39991;

DT 01-FEB-1995 (REL. 31, CREATED)

DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)

DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)

DE HYPOTHETICAL 136.1 KD PROTEIN IN YMA3-RIP1 INTERGENIC REGION.

GN YEL025C.

OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCAROMYCETALES;

OC SACCAROMYCETACEAE; SACCAROMYCES.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-S288C / AB972;

RA DIETRICH F.S., MOLLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,

RA AVILES E., BERNIO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,

RA CHUNG E., DUNCAN M., GUZMAN E., HARTZELL G., HONICKE-SMITH S.,

RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.

RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGREN R., OFFNER P., OH C.,

RA PEREL F.X., ROBERTS D., SEHL F., SCHRAMM S., SHOGREN T., SMITH V.,

RA TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.;

RA SUBMITTED (DEC-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U18530; G602392; -

KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.

FT TRANSMEM 73 93 POTENTIAL.

FT TRANSMEM 878 898 POTENTIAL.

FT TRANSMEM 1089 1109 POTENTIAL.

SO SEQUENCE 1188 AA; 136068 MW; 80AD1745 CRC32;

alignment_scores:

Quality: 61.00 Length: 63
 Ratio: 1.649 Gaps: 3
 Percent Similarity: 58.730 Percent Identity: 33.333

alignment_block:

US-09-030-606-224/rev x YEC5_YEAST ..

Align seg 1/1 to: YEC5_YEAST from: 1 to: 1188

```

253 GACGGCCACATTGTTGTCCTTCAGTATCAAGACACCTTTAGGGCTTCA 204
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:
729 AsperArgIleSerIlePheGlnPheGlySerThrPheTyrArgGluG1 745
203 TTTTGGAAAGTAGTATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 154
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:
745 yPheThrLysGlnMetThrSerLeuLysProGlnLeuGlnAsnHisIleP 762
153 AATATGCGCTTATACCTTTAATCATGATCATGATCATGATGATGG 104
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:
762 heTyrLeuProSerIle.....Pro.ValGluLeu..... 771
103 GAGTAAGGGGTACACACTCCCTACTACATGTCACAA 67
    |||||:||||:||||:||||:||||:||||:||||:||||:||||:
772 .....TyrThrValProAsnAsnLeuAsnGln 780

```

seq_name: SwissProt_37:DD31_DICDI

seq_documentation_block:

ID DD31_DICDI STANDARD; PRT; 269 AA.

AC Q02465;

DT 01-JUL-1993 (REL. 26, CREATED)

DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)

DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)

DE PRESPORE PROTEIN DD31.

GN SPIA.

OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).

OC EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-AX4;

RA MEDLINE; 91184531.

RA RICHARDSON D.L., HONG C.B., LOOMIS W.F.;

RA "A prespore gene, Dd31, expressed during culmination of Dictyostelium

RL discoidium.";

RL DEV. BIOL. 144:269-280(1991).

RN [2]

RP CHARACTERIZATION.

RX MEDLINE; 92275349.

RA RICHARDSON D.L., LOOMIS W.F.;

RA "Disruption of the sporulation-specific gene spia in Dictyostelium

RL discoidium leads to spore instability.";

RL GENES DEV. 6:1058-1070(1992).

CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH THE INNER FACE OF SPORE

CC COAT.

CC -1- DEVELOPMENTAL STAGE: EXPRESSED IMMEDIATELY PRIOR TO SPORE

CC FORMATION AND IN SPORES DURING CULMINATION.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X54452; G7164; -

DR PIR; S16671; S16671.

DR DICTYDB; DD04007; SPIA

SO SEQUENCE 269 AA; 30620 MW; 55B3C183 CRC32;

alignment_scores:

Quality: 60.00 Length: 109
 Ratio: 1.154 Gaps: 7
 Percent Similarity: 47.706 Percent Identity: 29.358

alignment_block:

US-09-030-606-224/rev x DD31_DICDI ..

Align seg 1/1 to: DD31_DICDI from: 1 to: 269

```

290 TTACCGTGTCCAGATGCACTTAAGAGGATGCGCCACATTT 241
    :::::::::::::: :::::::::::::: ::::::::::::::
72 VALPQALAPROIALthralaHisProArglnalaAspPheSerse 88
    :::::::::::::: :::::::::::::: ::::::::::::::
240 GTGTCCTTCAGATCAAGACCTTAAGGGCTTCATTTCAGAAAGTA 191
    ||| :::::::::::::: ::::::::::::::
88 rTyPromeIleArgLeuAsnLeu.....GlutGluA 99
    ||| :::::::::::::: ::::::::::::::
190 GATATTTCTCTCTCTCTCTTTAGACATATGATAATATGCCCCTTY 141
    :::::::::::::: :::::::::::::: ::::::::::::::
99 spIle.....AlaIleArgIuTyArgIlnIleValIysPhe 111
    :::::::::::::: :::::::::::::: ::::::::::::::
140 TATCCTTAACCATGATCATTTTCTTTGATGGGAGTACAG..... 96
    ||| ||| ||| ||| ||| |||
112 .....GlyIlePheValPheLeuTrpGluAlaAlaAla 123
    :::::::::::::: :::::::::::::: ::::::::::::::
95 .....GGTACACACCTCCCTACATACATGTC 71
    ||||| ||||| |||||
123 uValTyAsnTrpValValSerIleGlyThr.IleValTySerAlaVal 139
    :::::::::::::: :::::::::::::: ::::::::::::::
70 ACAACCTTTTCTTTCTTTCTTTCTTCTTCTTCTTCTTCTTCTT 30
    ||||| ||||| ||||| ||||| ||||| |||||
140 AspaSnpPheLeuAlaLeuPheTyMetIleValGlyValProthIle 156
    :::::::::::::: :::::::::::::: ::::::::::::::
29 ACTATTTTCTTAACAAGACCTT 7
    :::::::::::::: :::::::::::::: ::::::::::::::
156 u.TyrPheLeuThrArgIysLeu 163
    :::::::::::::: :::::::::::::: ::::::::::::::

```

seq_name: SwissProt_37:GYRB_MYCCA

seq_documentation_block:

```

ID GYRB_MYCCA STANDARD; PRT; 643 AA.
AC P50028;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE DNA GYRASE SUBUNIT B (EC 5.99.1.3).
GN GYRB.
OS MYCOPLASMA CAPRICOLUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
OC CAPRICOLUM GROUP.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 27343;
RX MEDLINE: 95129856.
RA SANO K.-I., MIYATA M.;
RT "the gyrb gene lies opposite from the replication origin on the
RT circular chromosome of Mycoplasma capricolum.";
RL GENE 151:181-183(1994).
CC -1- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
CC STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE
CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA.
CC -1- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
CC BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE
CC ENZYME FORMS AN A2B2 TETRAMER.
CC -1- SIMILARITY: BELONGS TO TYPE II TOPOISOMERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D28808; G938029; -
CC EMBL: D21231; G533332; -

```

```

DR EMBL: D26016; G532280; -
DR PROSITE: PS00177; TOPOISOMERASE_II; 1.
DR PRAM: PF00204; DNA_topoisomII; 1.
DR PRAM: PF00986; DNA_gyraseLC; 1.
DR HSSP: P06982; IAB6.
KW TOPOISOMERASE; ISOMERASE; ATP-BINDING.
SQ SEQUENCE 643 AA; 72333 MW; 093FD91D CRC32;

```

alignment_scores:

Quality:	Ratio:	Length:
60.00	1.622	72
Percent Similarity: 51.389	Percent Identity: 30.556	

alignment_block:

US-09-030-606-224 x GYRB_MYCCA ..

Align seg 1/1 to: GYRB_MYCCA from: 1 to: 643

```

43 ATAGAAACAACAAAAGAAAGTTGTGACATGTGTAGGAGTGTGT 92
    ||||| :::::::::::::: ::::::::::::::
29 IlleGlySerThrAspaSnpLysGlyLeuHisIleuValTrpGluIleVa 45
    :::::::::::::: :::::::::::::: ::::::::::::::
93 ACCCCTTCTCCCATCAAAAAAATGATGATGATGATGATGATGAT 142
    ||| :::::::::::::: ::::::::::::::
45 LaspaSnpAlaIleAspGlnAlaLeuAlaGlyTyrcysThrGln..... 59
    :::::::::::::: :::::::::::::: ::::::::::::::
143 AGGCGAATATTTTATCATATGCTCTTAAGAGAGAGAGAGAAATACTA 192
    ||||| :::::::::::::: ::::::::::::::
60 .....IleAspValIleLeuGluIuLysAspaSnpSerIleThr 71
    :::::::::::::: :::::::::::::: ::::::::::::::
193 CTTCCTCAATGATGAGACCCCTTAAGAGCTTGTGATACTGAAGACACA 242
    :::::::::::::: :::::::::::::: ::::::::::::::
72 ValIleAspaSnpGly.....ArgGlyIleProThrGlyMetHisLys 85
    :::::::::::::: :::::::::::::: ::::::::::::::
243 ATGTGGCCGCTCCATCC 258
    ||| :::::::::::::: ::::::::::::::
85 sThrGlyLysProThr 90
    :::::::::::::: :::::::::::::: ::::::::::::::

```

seq_name: SwissProt_37:CD14_RABIT

seq_documentation_block:

```

ID CD14_RABIT STANDARD; PRT; 372 AA.
AC Q28680;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE MONOCYTE DIFFERENTIATION ANTIGEN CD14 PRECURSOR.
GN CD14.
OS ORYCTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92268819.
RA LEE J.D., KATO K., TOBIAS P.S., KIRKLAND T.N., ULEVITCH R.J.;
RT "transfection of CD14 into 70Z/3 cells dramatically enhances the
RT sensitivity to complexes of lipopolysaccharide (LPS) and LPS binding
RT protein.";
RL J. EXP. MED. 175:1697-1705(1992).
CC [2]
CC SEQUENCE FROM N.A.
CC TISSUE-MYELOID;
CC RA ISHIDA T., SETOGUCHI M., MATSURA K., YASUNORI H., AKIZUKI S.,
CC YAMAOKA S.;
CC SUBMITTED (AUG-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -1- FUNCTION: SERVES AS AN LPS RECEPTOR CONTROLLING CELL ACTIVATION
CC UNDER PHYSIOLOGICAL CONDITIONS. WHEN LPS BINDS TO CD14 THE CELLS
CC BECOME ACTIVATED AND RELEASE CYTOKINES AND UP-REGULATE CELL
CC SURFACE MOLECULES, INCLUDING ADHESION MOLECULES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

RT "Reconstructing the taxonomic radiation of dasyurine marsu-
DM (invertebrates) 135 2000 and extinction of 2000 2000"

This Page Blank (uspto)

OC Potamoctrygon.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MARTIN A.P., LOVEJOY N., BERMINGHAM E.;
 RT "Marine incursion into South America."
 RL Nature 396:421-422(1998).
 DR EMBL; AF110625; AAD12291.1; -
 KW Mitochondrion.
 FT NON_TER 1 1
 FT NON_TER 255 255
 SQ SEQUENCE 255 AA; 29047 MW; E8F7E237 CRC32;

alignment_scores:
 Quality: 71.50 Length: 84
 Ratio: 1.589 Gaps: 4
 Percent Similarity: 53.571 Percent Identity: 30.952

alignment_block:
 US-09-030-606-224/rev x 099907 ..

Align seg 1/1 to: 099907 from: 1 to: 255

```

188 ATTTCTCTCTCTCTCT.....TTAGAAC 163
   |||:|||||:|||||:|||||:|||||:
121 lletpelyglypheserValaspasnaIathrleuthrargpetherh 137
   |||:|||||:|||||:|||||:|||||:
162 ATATGATAAATATTGCCCTTAT.....CCTTAACCATGATCCAT 119
   |||:|||||:|||||:|||||:|||||:
137 rphenIsphleupheprpophmetIlealalaIeuthmet.IleHis 153
   |||:|||||:|||||:|||||:|||||:
118 TTTTGTGATGGGAGTAAGGGTACACATCC..... 83
   |||:|||||:|||||:|||||:|||||:
154 LeuendheupheuhIsGluAlaGlySerAsnAsnProthrGlyLeuthrSe 170
   |||:|||||:|||||:|||||:|||||:
82 .....TACTACATGTCA 70
   |||:|||||:|||||:|||||:|||||:
170 rAspMetAspLysIleGlnpheHisProtyrTyrrTrrpTyLysAspLeuV 187
   |||:|||||:|||||:|||||:|||||:
69 CAACCTTTTCTTTGTGTCCTATGTTGTAACGTACTATTCT 20
   |||:|||||:|||||:|||||:|||||:
187 alGlyPhepHeuLeuLeuLeuLeuAlaIleuAlaLeuPhehr 203

```

seq_name: sp_organelle:099906

seq_documentation_block:

ID 099906 PRELIMINARY; PRT; 210 AA.
 AC 099906;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, last annotation update)
 DE CYTOCHROME B (FRAGMENT).
 OS Potamoctrygon orbignyi.
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Rajiformes; Myliobatoidel; Potamoctrygonidae;
 OC Potamoctrygon.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MARTIN A.P., LOVEJOY N., BERMINGHAM E.;
 RT "Marine incursion into South America."
 RL Nature 396:421-422(1998).
 DR EMBL; AF110625; AAD12290.1; -
 KW Mitochondrion.
 FT NON_TER 1 1
 FT NON_TER 210 210
 FT NON_TER 23926 MW; 7235786B CRC32;
 SQ SEQUENCE 210 AA; 23926 MW; 7235786B CRC32;

alignment_scores:
 Quality: 70.50 Length: 83
 Ratio: 1.602 Gaps: 4
 Percent Similarity: 53.012 Percent Identity: 31.325

alignment_block:
 US-09-030-606-224/rev x 099906 ..

Align seg 1/1 to: 099906 from: 1 to: 210

```

188 ATTTCTCTCTCTCTCT.....TTAGAAC 163
   |||:|||||:|||||:|||||:|||||:
121 lletpelyglypheserValaspasnaIathrleuthrargpetherh 137
   |||:|||||:|||||:|||||:|||||:
162 ATATGATAAATATTGCCCTTAT.....CCTTAACCATGATCCAT 119
   |||:|||||:|||||:|||||:|||||:
137 rphenIsphleupheprpophmetIlealalaIeuthmet.IleHis 153
   |||:|||||:|||||:|||||:|||||:
118 TTTTGTGATGGGAGTAAGGGTACACATCC..... 83
   |||:|||||:|||||:|||||:|||||:
154 LeuendheupheuhIsGluAlaGlySerAsnAsnProthrGlyLeuthrSe 170
   |||:|||||:|||||:|||||:|||||:
82 .....TACTACATGTCA 70
   |||:|||||:|||||:|||||:|||||:
170 rAspMetAspLysIleGlnpheHisProtyrTyrrTrrpTyLysAspLeuV 187
   |||:|||||:|||||:|||||:|||||:
69 CAACCTTTTCTTTGTGTCCTATGTTGTAACGTACTATT 23
   |||:|||||:|||||:|||||:|||||:
187 alGlyPhepHeuLeuLeuLeuLeuAlaIleuAlaLeuPhe 202

```

seq_name: sp_organelle:Q34293

seq_documentation_block:

ID Q34293 PRELIMINARY; PRT; 244 AA.
 AC Q34293;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
 DT 01-JAN-1999 (TREMBlrel. 09, last annotation update)
 DE CYTOCHROME B (FRAGMENT).
 OS Dasysyllis akajel.
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Rajiformes; Myliobatoidel; Dasysyllidae; Dasysyllis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA KITAMURA T., TAKEMURA A., WATABE S., TANUCHI T., SHIMIZU M.;
 RT "Molecular phylogeny of the sharks and rays of superorder squalea
 based on mitochondrial cytochrome b gene."
 RL Fisheries Sci. 62:340-343(1996).
 DR EMBL; D50027; BAA08754.1; -
 DR PFAM; PF00032; cytochrome_b.C; 1.
 DR PFAM; PF00033; cytochrome_b.N; 1.
 KW Mitochondrion.
 FT NON_TER 1 1
 FT NON_TER 244 244
 FT NON_TER 27814 MW; EBBB32CE CRC32;
 SQ SEQUENCE 244 AA; 27814 MW; EBBB32CE CRC32;

alignment_scores:
 Quality: 70.50 Length: 102
 Ratio: 1.282 Gaps: 5
 Percent Similarity: 53.922 Percent Identity: 28.431

alignment_block:

US-09-030-606-224/rev x Q34293 ..
 Align seg 1/1 to: Q34293 from: 1 to: 244

```

230 GATCAAGCACCTTTAAGGCTTCATTYYGAAAGATAGT..... 189
   |||:|||||:|||||:|||||:|||||:
12 lIethrsnLeuLeuSerAlaIeupProtyrIleGlyAspThrLeuValGl 28

```

```

188 ..ATTTCCTGCTCCCTCT..TTA 167
|||||.....|
28 ntprlletrglycylphresertleasrsalatlrltuthrathrpher 45
|||||.....|
45 hehrghenhsrheuleurpherorphevalleallaleuthmet.il 61
|||||.....|
122 CCATTTTTTGTGGAGGAGTAAGGCGTAGACACTGCC.....83
|||||.....|
61 enisleuleurpheullehnglturhglsesarsaanprotthrglyleurt 78
|||||.....|
82 .....TACTGCAT 74
|||||.....|
78 hserasnhrasryslleprophenisProtygrthethtyLysasp 94
|||||.....|
73 GTCACAACCTTTCTTTTGTTGTTCCATTTGTTGTTGAACCTAACTAT 24
|||||.....|
95 levalvGlYrheherleleuleurpheullethleuleualleu 111
|||||.....|
23 tttct 20
|||||.....|
111 ethr 112

seq_name: sp_organelle:099908

seq_documentation_block:
ID_099908 PRELIMINARY; PRT; 252 AA.
AC 099908:
DT 01-MAY-1999 (TREMBLEl. 10, Created)
DT 01-MAY-1999 (TREMBLEl. 10, Last sequence update)
DT 01-MAY-1999 (TREMBLEl. 10, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
OS Potamoitrygon castexi.
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes; Elasmobranchii; Rajiformes; Myliobatoidae; Potamoitrygonidae; OC Potamoitrygon.
RN (1)
RP SEQUENCE FROM N.A.
RA MARTIN A.P., LOVEJOY N., BERMINGHAM E.; RT "Marine Incursion Into South America."; RL Nature 396:421-422(1998). DR EMBL; AF110627; AAD1292.1; -. KW Mitochondrion. FT NON_TER 1 1 FT NON_TER 252 252 SQ SEQUENCE 252 AA; 28712 MW; 353314BE CRC32;

alignment_scores:
Quality: 70.50 Length: 83
Ratio: 1.602 Gaps: 4
Percent Similarity: 53.012 Percent Identity: 31.325

alignment_block:
US-09-030-606-224/rev x 099908 ..
Align seg 1/1 to: 099908 from: 1 to: 252

188 ATTTCCTGCTCCCTCT...TTAGAAC 163
|||||.....|
118 lletrglycylrheherleleuasrsalatlrltuthrathrpher 134
|||||.....|
162 ATATGATAAATATGGCCTTAT....CCTTAACCATATACAT 119
|||||.....|
134 rphenhsrheuleurpherorphetlealaaleuthmet.llehis 150
|||||.....|
118 TTTTTTGTGGAGGAGTAAGGCGTAGACACTGCC.....83
|||||.....|
151 leuleurpheullestgluaaglysearsaasnProtthrglyleuthrse 167
|||||.....|
TACTGCATGTCA 70
```

```

167 rnsphetslpysllieginpnehisprotyrtytrpttyllysaspleu 184
168 CAAACCTTTTCTTTTGTGTGTCCTATTGGTTGTAAGTACTATTT 23
169 ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 aiclypsherpheleuendeuleuenaialaleuAialeuphe 199

seq_name: sp_organelle:09919

seq_documentation_block:
ID 09919 PRELIMINARY; PRT; 255 AA.
AC 09919;
DT 01-MAY-1999 (TREMBLE). 10, Created)
DT 01-MAY-1999 (TREMBLE). 10, Last sequence update)
DT 01-MAY-1999 (TREMBLE). 10, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
OS Hmanctura pacifica.
OC Mitochondrion.
OC Eukaryota. Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes.
OC Elasmobranchii; Rajiformes; Myliobatoidei; Dasypatiidae; Hmanctura.
RN [1]
RP SEQUENCE FROM N.A.
RA MARTIN A.P., LOVEJOY N., BERMINGHAM E.;
RT "Marine incursion into South America.";
RL Nature 396:421-422(1998).
DR EMBL; AF110638; AADI2303.1; -.
KW Mitochondrion.
FT NON_TER 1 1
FT TER 255 255
FT FT
SQ SEQUENCE 255 AA; 28765 MW; 46159853 CRC32;

alignment_scores:
Quality: 70.50 Length: 102
Ratio: 1.306 Gaps: 5
Percent Similarity: 52.941 Percent Identity: 27.451

alignment_block:
US-09-030-606-224/rev x 09919 ..

Align seg 1/1 to: 09919 from: 1 to: 255

230 GATATCAAGACACCTTTAAGGCGTCCATTGAGACAAAGTACT..... 189
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 IletnrasleuleusealaleuProtyrileglyasprthleuValgl 119
188 ... ATTTCCTCTCTCTCT..... TTAA 167
119 n**ValtrpglyglypserseValaspsanAlthleuthGlnphee 136
166 GAACATATGATAAATATTTGCCCTTY.....TATCCTTTAACATGTAT 123
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
136 hehtrpnehisheleuherProprheValilevalAlaleuthMet.11 152
122 CCAATTTTGTGATGGGAGTAAGGGGTACACACTCC..... 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
152 ehlsleuLeupheleuhsGlnalaglyserAsnAsnProthrglyleu 169
82 .....TACTAAT 74
169 laseAsnAlthAspysllieGlnpnehisProtyrtyrserTyrllysgln 185
73 GTCACAAACCTTTTCTTTGTGTGTCCTATTGGTTGTAAGTACTATTT 24
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
186 leuLeuGlypsherpheleuendeuleuenaialaleuAialeuph 202
23 TTCT 20
:::
1:::
202 ethr 203

seq_name: sp_organelle:09920

seq_documentation_block:

```

```
seq_name: sp_oranelle:099920
seq_documentation_block:
```

ID	009920	PRELIMINARY;	PRT;	255	AA.
AC	009920;				
DT	01-MAY-1999	(TREMBLER).	10,	Created)	
DT	01-MAY-1999	(TREMBLER).	10,	Last sequence update)	
DT	01-MAY-1999	(TREMBLER).	10,	Last annotation update)	
DE	CYTOCROME B (FRAGMENT).				
OS	Himantura scymnaeae.				
OG	Mitochondrion.				
OC	Eulachyonidae.				
OC	Elasmobranchii; Rajiformes; Myliobatoidae; Dasysyllidae; Himantura.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	MARTIN A.P., LOVEJOY N., BERINGHAM E.;				
RT	"Marine Incursion into South America.";				
RL	Nature 396:421-422(1998).				
DR	EMBL: AF110639; AAD12304.1; -.				
KW	Mitochondrion.				
FT	NON_TER	1			
FT	NON_TER	255	1		
Q0	SEQUENCE	255	255	MM;	453B4FF4 CRC32;
		255	AA;	28904	

```

alignment_scores:
  quality: 70.50
  ratio: 1.602
  percent_similarity: 52.381
  length: 84
  gaps: 4
  percent_identity: 30.952

```

alignment_block:

US-09-030-606-224/rev x 099920 .

Align seg 1/1 to: 099920 from: 1 to: 255

[illegible]

	seq_documentation_block:	
ID	099904	PRELIMINARY; PRT; 255 AA.
AC	099904;	
DT	01-MAY-1999 (TREMBLE)	10, Created
DT	01-MAY-1999 (TREMBLE)	10, Last sequence update
DT	01-MAY-1999 (TREMBLE)	10, Last annotation update
DE	CYTOCHROME B (FRAGMENT).	
OS	Dasyatis hawaiiensis.	
OG	Mitochondrion.	
OC	Euarystia Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes	
CC	Elasmobranchii; Rajiformes; Myliobatoidae; Dasyatidae; Dasyatis.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	MARTIN A.P., LOVEJOY N., BERMINGHAM E.;	
RL	"Marine incursion into South America."	
RT	Nature 396:421-422(1998).	
DR	EMBL; AF110623; AAD12288.1; -.	
FW	Mitochondrion.	
NON_TER	1	1

FT	NON_TER	255	255
SQ	SEQUENCE	255 AA; 28769 MW;	2A435F7B CRC32;

alignment_scores:		
Quality:	69.50	Length: 84
Ratio:	1.544	Gaps: 4
Percent Similarity:	53.571	Percent Identity: 30.952

```
alignment_block:
```

US-09-030-606-224/rev x 099904

Align seg 1/1 to: 099904 from: 1 to: 255

[illegible]

```

seq_documentation_block:
ID      021173          PRELIMINARY;          PRT;          378 AA.
AC      021173;
DT      01-JAN-1998 (TREMblrel. 05, Created)
DR      01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT      01-MAY-1999 (TREMblrel. 10, Last annotation update)
DE      CYTOCHROME-B.
OS      Heliothobius argenteocinctus (Silvery mole-rat).
OG      Mitochondrion.
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
        Eutheria; Rodentia; Hystricognathi; Bathyrigidae; Heliothobius.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 97370639.
RA      FAULKES C.G., ABBOTT D.H., O'BRIEN H.P., LAU L., ROY M.R., WAYNE R.K.,
        BRUFORD M.W.;
RT      "Micro- and macrogeographical genetic structure of colonies of naked
        mole-rats Heterocephalus glaber.";
RL      Mol. Ecol. 6:615-628 (1997).
CC      -I- CATALYTIC ACTIVITY: OH(2) + 2 FERRICYTOCHROME C = Q + 2
        FERROCYTOCHROME C.
CC      -I- COFACTOR: TWO HEME GROUPS
CC      (B562 AND B566) WHICH ARE NOT COVALENTLY BOUND TO THE PROTEIN
        (BY SIMILARITY).
CC      EMBL; U87527; AAC53270.1; -.
DR      PFMW; PF00032; cytochrome_b.C; 1.
DR      PFMW; PF00033; cytochrome_b.N; 1.
KW      Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
        Heme.
SQ      SEQUENCE 378 AA; 42889 MW; C32032CF CRC32;

```

```

alignment_scores:
  Quality: 69.50
  Ratio: 1.264
  Percent Similarity: 53.922
  Length: 1022
  Gaps: 6
  Percent Identity: 29.412

```

alignment_block:

US-09-030-606-224/rev x 021173 ..

Align seg 1/1 to: 021173 from: 1 to: 378

```

230 GTATCAAGACACCTTTAAGGCTTCATTTTGAGAAAGTACT..... 189
      ::::::::::::::::::::|
145 IlerhansgnleuseralaleprotyvalglyserSerleuvalgl 161
      ::::::::::::::::::::|
188 ... ATTTCCTCTCCTCTCT.....TTTA 167
      |||:::|
161 utripilrtpglypserValasplyserThrleuTharpgpner 178
      ::::::::::::::::::::|
166 GACATATGATAAATATATGCTTAT.....CCTTAACCATGAT 123
      ::::::::::::::::::::|
178 healaphnehisphelleuProphellellelealaleuThmet.Va 194
      ::::::::::::::::::::|
122 CCATTTTGTGATGGGAGTAAAGGCTACACACT..... 86
      ::::::::::::::::::::|
194 IHisleuPhelHeuHisglutThrGlySerAsnAsnProserglyleu 211
      ::::::::::::::::::::|
85 .....CCTTACTAC.....AAT 74
      ::::::::::::::::::::|
211 spSerAsnSerAspLysIleProphHisProtyTyrSerPhelysAT 227
      ::::::::::::::::::::|
73 GTCAAAACCTTTTCTTTGTTGCTTATGTTGTTGTTGTTGTTGTT 24
      ::::::::::::::::::::|
228 leuLeuglypneMetleuLeuValleuLeuSerleuValleuPh 244
      ::::::::::::::::::::|
23 TTTCT 20
      ::::::::::::::::::::|
244 ethr 245

```

seq_name: sp_oranelle:Q34820

seq_documentation_block:

```

ID Q34820; PRELIMINARY; PRT; 266 AA.
AC Q34820;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DE 01-MAY-1999 (TREMBLrel. 10, last annotation update)
DE CYTOCHROME B (FRAGMENT).
OS Isothrix pagurus.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Hystriocognathi; Echimyidae; Isochirix.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-INPA 2463.
RX MEDLINE: 94184505.
RA DA SILVA M.N., PATTON J.L.;
RT "Amazonian phylogeography: mtDNA sequence variation in arboreal
RT echimyid rodents (Caviomorpha)".
RL Mol. Phylogenet. Evol. 2:243-255(1993).
CC -1- CATALYTIC ACTIVITY: OH(2) + 2 FERRICYTOCHROME C - Q + 2
CC FERROCYTOCHROME C.
CC -1- COFACTOR: TWO HEME GROUPS
CC (B562 AND B566) WHICH ARE NOT COVALENTLY BOUND TO THE PROTEIN
CC (BY SIMILARITY).
CC EMBL: L23348; AAA31858.1; -.
DR PPM: PF000033; cytochrome b.N: 1.
KW Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
KW Heme.
FT NON_TER 266
FT SEQUENCE 266 AA; 30088 MW; 2F76B588 CRC32;

```

alignment_scores:

Quality: 68.50 Length: 83
 Ratio: 1.712 Gaps: 4
 Percent Similarity: 48.193 Percent Identity: 28.916

seq_documentation_block:

US-09-030-606-224/rev x Q34820 ..

Align seg 1/1 to: Q34820 from: 1 to: 266

```

217 TTTAAGGCTTCATTTTGAGAAAGTATTTTCTCTCTCTCTTT 168
      |||:::|
178 PhehealaphneHis.....PheleleuProphelleil 189
      ::::::::::::::::::::|
167 AGACATATGATAAATATATGCTTATGCTTATACATGTTGCAT 118
      ::::::::::::::::::::|
189 ethalaleMetVal.....MetleHis 197
      ::::::::::::::::::::|
117 TTTTGTGATGGGAGTAAAGGCTACACACT..... 86
      ::::::::::::::::::::|
197 euleuPhelHeuHisglutThrGlySerAsnAsnProserglyleuAsnSer 213
      ::::::::::::::::::::|
85 .....CCTTACTAC.....AATGTCAC 69
      ::::::::::::::::::::|
214 AsperAspLysIleProphHisProtyTyrThrIleLysAspIle 230
      ::::::::::::::::::::|
68 AAACCTTTTCTTTGTTGCTTATGTTGTTGTTGTTGTTGTT 20
      ::::::::::::::::::::|
230 uGlyPheMetPheMetleuLeuLeuValleuValleuPheSer 246
      ::::::::::::::::::::|

```

seq_name: sp_vertibrate:057604

seq_documentation_block:

```

ID 057604; PRELIMINARY; PRT; 571 AA.
AC 057604;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, last annotation update)
DE THROMBOWCIN PRECURSOR (FRAGMENT).
DE Gallus gallus (Chicken).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RA MCNAGN K.M., PETERSSON I., ROSSI F., FLAMME I., SHEVCHENKO A.,
RA WANN M., GRAF T.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: Y13978; CAA74311.1; -.
KW Signal.
KW SIGNAL.
FT NON_TER 1 19 POTENTIAL.
FT SEQUENCE 571 AA; 61014 MW; 9FBC76E5 CRC32;

```

alignment_scores:

Quality: 67.00 Length: 46
 Ratio: 2.310 Gaps: 1
 Percent Similarity: 63.043 Percent Identity: 30.435

alignment_block:

US-09-030-606-224 x 057604 ..

Align seg 1/1 to: 057604 from: 1 to: 571

```

59 GAAAAGTTTGTGATGAGTAAAGGAGTGTACCCCTTACTCCCAT 108
      |||:::|
360 gluVallelcygsluAspGlnlleGlyValalargProilleuAsnle 376
      ::::::::::::::::::::|
109 CAAA.....AAAAAATGATACATGTTTAAAGGATARAAGGCAATAT 152
      ::::::::::::::::::::|
376 ulysgluGluLysThrCysAspAspTrpLysAlaSerAsnGluAlap 393
      ::::::::::::::::::::|
153 TTTATCATATGTTCTTAAAGGAGAGGAGGAGGAGGAGGAGGAGG 190
      ::::::::::::::::::::|
393 hepneGluValPheCysSerGlyAlrGArGHisAlapne 405
      ::::::::::::::::::::|

```

seq_name: sp_oranelle:099912

seq_documentation_block:

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:29:54; Search time 1809.22 Seconds

(without alignments)
348.886 Million cell updates/sec

US-09-030-606-224

Title: Perfect score: 320
Sequence: 1 CCCCTGAAGCTTCTCTTA.....TTTAACTCMGCAATGTGAC 320

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database:

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: em_est20: *
21: em_est21: *
22: em_est22: *
23: em_est23: *
24: em_est24: *
25: em_est25: *
26: em_est26: *
27: em_est27: *
28: em_est28: *
29: em_est29: *
30: em_est30: *
31: em_est31: *
32: em_est32: *
33: em_est33: *
34: em_est34: *
35: em_est35: *
36: em_est36: *
37: em_est37: *
38: em_est38: *
39: em_est39: *
40: em_est40: *
41: em_est41: *
42: em_est42: *
43: em_est43: *
44: em_est44: *
45: em_est45: *
46: em_est46: *
47: em_est47: *
48: em_est48: *
49: em_est49: *
50: em_est50: *
51: em_est51: *
52: em_est52: *
53: em_est53: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	202.2	63.2	512	21	R20590	R20590 y949d06.r1
2	201.2	62.9	285	40	245397	245397 HSC20D041.n
3	165.4	51.7	342	32	AA334754	AA334754 EST39036
4	152	47.5	177	25	N56292	N56292 J5669F.Hum
5	136.8	42.8	315	32	AA373221	AA373221 EST85227
6	40.4	12.6	214	43	A0033864	A0033864 A0033864
7	38.4	12.0	700	48	A1557583	A1557583 PC2.1-1.D
8	37.2	11.6	310	40	C93124	C93124 C93124 Dict
9	36.8	11.5	249	41	A1054289	A1054289 q176604.x
10	36.8	11.5	269	42	A1144030	A1144030 q163808.x
11	36.6	11.4	496	23	H17024	H17024 ym39d10.s1
12	36.6	11.4	434	23	L47084	L47084 SCMRAP171.C
13	36.6	11.4	300	35	C57199	C57199 C57199 Yuj1
14	36.6	11.4	300	36	C57822	C57822 C57822 Yuj1
15	36	11.2	300	36	C59863	C59863 C59863 Yuj1
16	35.8	11.2	354	44	A0038221	A0038221 A0038221
17	35.2	11.0	504	35	AA550294	AA550294 1423m3.gm
18	35	10.9	587	50	A1667004	A1667004 tc24d11.y
19	34.8	10.9	471	46	A1404883	A1404883 GH24733.S
20	34.6	10.8	432	36	AA604491	AA604491 no73809.s
21	34.4	10.8	305	37	AA721414	AA721414 n274b11.s
22	34.4	10.8	520	39	AA825604	AA825604 cd57n06.s
23	34.4	10.8	312	43	A1239707	A1239707 qn43b02.x
24	34.4	10.8	568	44	A0039059	A0039059 A0039059
25	34.4	10.8	384	49	A1630952	A1630952 ty99g12.x
26	34.2	10.7	442	26	W47600	W47600 zc35g07.r1
27	34.2	10.7	323	34	C24351	C24351 C24351 Dict
28	34.2	10.7	171	40	C90553	C90553 C90553 Dict
29	34.2	10.7	412	43	A1178240	A1178240 EST221905
30	34.2	10.7	406	51	A1705892	A1705892 UT-R-AC1
31	34	10.6	533	34	C24372	C24372 C24372 Dict
32	34	10.6	699	45	A0001859	A0001859 A0001859
33	34	10.6	612	47	A1525038	A1525038 promrna-5
34	34	10.6	373	48	A1582163	A1582163 tg63b09.x
35	34	10.6	477	53	HSM000181	A1035868 Homo sap1
36	33.8	10.6	464	20	Z32876	Z32876 HHEX1N.Alt
37	33.8	10.6	466	22	R22059	R22059 yb25g02.s1
38	33.8	10.6	583	25	N50519	N50519 yb89b04.s1
39	33.8	10.6	388	25	N97950	N97950 1711C3 czap
40	33.8	10.6	383	25	N97969	N97969 1711C3 czap
41	33.8	10.6	427	31	AA279549	AA279549 z88b09.r1
42	33.8	10.6	565	46	A1407578	A1407578 EST35868
43	33.8	10.6	353	47	A1522468	A1522468 fb19b08.x
44	33.6	10.6	312	49	AV020572	AV020572 AV020572
45	33.6	10.5	602	34	AA532223	AA532223 CPEST.501

ALIGNMENTS

RESULT 1
R20590
LOCUS
DEFINITION y949d06.r1 Soares Infant brain 1n1B Homo sapiens cDNA clone
IMAGE:36244 5', mRNA sequence.
ACCESSION R20590
NID 9775371
VERSION R20590.1 GI:775371

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS 1 (bases 1 to 512)
 Hillier, L., Clark, N., Dubuque, T., Elliston, R., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 High quality sequence stops: 321
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: M13RP1
 High quality sequence stop: 321.
FEATURES
 source
 1..512
 /organism="Homo sapiens"
 /db_xref="GDB:408745"
 /db_xref="taxon:9606"
 /clone_1ib="IMAGE:36244"
 /sex="female"
 /dev_stage="73 days post natal"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: whole brain; Vector: lafmid BA; Site: 1: Not I; Site: 2: Hind III; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15',
 A - GAGTGAAGAGCTTCTGTGTAAGAAATAGTACAGTTACACCAATAGAGACACAAAAGA 60
 76 CCCTGAGAGCTTCTGTGTAAGAAATAGTACAGTTACACCAATAGAGACACAAAAGA 135
 61 AAAAGTTTGACATGTATAGGAGTGTACCCCTTACCCCATCAAAAAAAT 120
 136 AAAAGTTTGACATGTATAGGAGTGTACCCCTTACCCCATCAAAAAAAT 195
 121 GGATACATGTTAAAGATTAAGAGGCAATATTTATCATATGTCTTAAAGAGAGAA 180
 196 GGATACATGTTAAAGATTAAGAGGCAATATTTATCATATGTCTTAAAGAGAGAA 255
 181 GAGAAATACTACTTCTCRAAA--TGAAGCCCTTAAAGTG 221
 256 GAGAAATACTACTTCTCRAAAATGGAAGCCCTTAAAGGG 298
BASE COUNT 141 a 93 c 111 g 158 t 9 others
ORIGIN
 Query Match 63.2%; Score 202.2; DB 21; Length 512;
 Best Local Similarity 96.9%; Pred. No. 7.8e-38;
 Matches 216; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

DEFINITION HSC20D041 normalized infant brain cDNA Homo sapiens cDNA clone
 c-20d04, mRNA sequence.
ACCESSION 245397
VERSION 9574610
KEYWORDS 245397.1 GI:574610
SOURCE EST.
ORGANISM human.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS 1 (bases 1 to 285)
 Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes, M.D., Duprat, S., Houlgatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F., Mitchell, H., Marigae-Samson, R., Pietu, G., Poullot, Y., Sebastiani-Kabakchis, C., and Tessier, A.
TITLE IMAGE: molecular integration of the analysis of the human genome and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407250.
FEATURES
 source
 1..285
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1ib="c-20d04"
 /clone_1ib="normalized infant brain cDNA"
 /sex="female"
 /tissue_type="total brain"
 /dev_stage="3 months old"
 /note="Organ: brain; Vector: lafmid BA; Site: 1: HindIII; Site: 2: NotI; sex: female; dev_stage=3 months old; isolate-muscular atrophy patient; tissue_type-total brain; total mRNA was oligo-(dT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA vector. Clone library from B.Souares, Psychiatry Dept. Columbia University, USA. Normalization method: Bento Soares, P.N.A.S. in press"
BASE COUNT 101 a 44 c 52 g 86 t 2 others
ORIGIN
 Query Match 62.9%; Score 201.2; DB 40; Length 285;
 Best Local Similarity 98.0%; Pred. No. 1.4e-37;
 Matches 200; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 1 CCCTGAGAGCTTCTGTGTAAGAAATAGTACAGTTACACCAATAGAGACACAAAAGA 60
 82 CCCTGAGAGCTTCTGTGTAAGAAATAGTACAGTTACACCAATAGAGACACAAAAGA 141
 61 AAAAGTTTGACATGTATAGGAGTGTACCCCTTACCCCATCAAAAAAAT 120
 142 AAAAGTTTGACATGTATAGGAGTGTACCCCTTACCCCATCAAAAAAAT 201
 121 GGATACATGTTAAAGATTAAGAGGCAATATTTATCATATGTCTTAAAGAGAGAA 180
 202 GGATACATGTTAAAGATTAAGAGGCAATATTTATCATATGTCTTAAAGAGAGAA 261
 181 GAGAAATACTACTTCTCRAAA 204
 262 GAGAAATACTACTTCTCRAAAAT 285

[illegible]

DB	121	CATGACTTGAGACAGCGTACTGTTGACATTTTACAGTACTGATGTGAC	169
RESULT	4		
N56292			
DEFINITION			
Accession	N56292	177 bp	EST
Version	J35699F	Human fetal heart, Lambda ZAP Express	20-FEB-1996
KeyWords	clone J35699 5', mRNA sequence.		
Source	N56292.1	GI:1199140	
Organism	human.		
Reference	Homo sapiens		
Authors	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
Title	Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
Journal	1 (bases 1 to 177)		
Comment	cdnas from fetal heart		
	unpublished (1995)		
	On May 18, 1995 this sequence version replaced gi:811148.		
FEATURES			
Source			
Contact: J1ew CC			
Department of Laboratory Medicine and Pathobiology			
University of Toronto			
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5			
Tel: 416/9788758			
Fax: 416/9785650			
Email: j1ewcc@utcc.utoronto.ca			
Seq primer: TCCCAAGATTCGGCAGCAG.			
Location/Qualifiers			
1. 177			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/map="11"			
/clone="J35699"			
/lab_host="E. coli XL1-Blue"			
/note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2: XhoI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XhoI-Oligo dT adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI for directional cloning into predigested Lambda ZAP Express."			
BASE COUNT	69 a	27 c	32 g
ORIGIN			
Query Match	47.58;	Score 152;	DB 25;
Best Local Similarity	92.96;	Pred. No. 3.9e-26;	Length 177;
Matches 158; Conservative	1;	Mismatches 11;	Indels 0;
		Gaps 0;	
QY	1	CCCCGAAGGCTTCTGTAGCAAAATAGTACAGTTACACCAATAGGACACAAAGA	60
Db	8	CCCCGAAGGCTTCTGTAGCAAAATAGTACAGTTACACCAATAGGACACAAAGA	67
QY	61	AAAAGTTTGACATTGTAGTAGGAGATGTGTACCCCTTACTCCCATCAAAAAAAT	120
Db	68	AAAAGTTTGACATTGTAGTAGGAGATGTGTACCCCTTACTCCCATCAAAAAAAT	127
QY	121	GGATACATGCTTAAAGATATAAGGCAATATTTATCATATGTCTTAA	170
Db	128	GGATACATGCTTAAAGATATAAGGCGCAATATTTATCATATGTCTTAA	177
RESULT	5		
LOCUS	AA373221	315 bp	mRNA
DEFINITION	EST85227	HSC172 cells 1	Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION	AA373221		
VERSION	92025541		
ORIGIN	AA373221.1	GI:2025541	

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 315)
Adams,M.D., Keriavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-ai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S.,
Glodet,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
Keller,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Uteback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dlake,D., Feng,D.-F., Ferris,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weisner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M., and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl.), 3-174 (1995)
96026280
JOURNAL On May 18, 1995 this sequence version replaced gi:811155.
MEDLINE Other-ESTs: THC168716
COMMENT Contact: Keriavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse
Location/Qualifiers
1. 315
/organism="Homo sapiens"
/db_xref="ATCC (lnhost):17582"
/db_xref="taxon:9606"
/clone_lib="HSC172 cells 1"
/cell_type="fibroblast"
/cell_line="HSC172 (16PDL)"
/dev_stage="fetal"
/note="Organ: lung; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
BASE COUNT 86 a 65 c 64 g 99 t 1 others
ORIGIN

AU033864
LOCUS 214 bp mRNA EST 28-APR-1999
DEFINITION AU033864 Dictyostelium discoidium SL (H.Urushihara) Dictyostelium
discoidium cDNA clone SLB538, mRNA sequence.
ACCESSION AU033864
MID 93799288
VERSION AU033864.1 GI:3799288
KEYWORDS EST.
SOURCE Dictyostelium discoidium.
ORGANISM Dictyostelium discoidium.
REFERENCE Eukaryota; Dictyostellida; Dictyostelium.
AUTHORS 1 (bases 1 to 214)
Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Watanabe,R., Mitra,B.N., Pt,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
The Dictyostelium development cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development
DNA Res. 5 (6), 335-340 (1998)
99156227
JOURNAL On May 19, 1998 this sequence version replaced gi:2846208.
MEDLINE
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan
Email: d402huesakura.cc.tsukuba.ac.jp
PROJECT - 'Dictyostelium discoidium cDNA project in Japan'
POLYA-No.
FEATURES
source
1. 214
/organism="Dictyostelium discoidium"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="SLB538"
/cell_line="Dictyostelium discoidium SL (H.Urushihara)"
/dev_stage="slug"
BASE COUNT 149 a 22 c 12 g 31 t
ORIGIN

Query Match 12.6%; Score 40.4; DB 43; Length 214;
Best Local Similarity 51.1%; Pred. No. 2.9;
Matches 92; Conservative 1; Mismatches 87; Indels 0; Gaps 0;
RESULTS 7
A1557583 700 bp mRNA EST 23-MAR-1999
LOCUS p12.1-1.D10.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
DEFINITION A1557583
ACCESSION A1557583
MID 94489946
VERSION A1557583.1 GI:4489946
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 700)
Huang,G.M., Ng,W., Faras,J., Chen,L., Liang,H.A., Gordon,D., Jun
Yu,J. and Hood,L.

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="8"
/clone IMAGE:1862406
/clone_1lb="NCI_CGAP_Ov26"
/sex="female"

```

```

/organism="homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1861142"
/clone_11b="NCI_CGAP_OV26"
/sex="female"
/tissue_type="papillary serous carcinoma"
/dev_host="adult"
/lab_host="DH10B"
/note="Organ: ovary; Vector: pAMP1; mRNA made from
papillary serous ovarian carcinoma, cDNA made by oligo-AT
priming. Directionally cloned. Site selected on agarose
gel, average insert size 500 bp. Primary library,
non-amplified."

```

```

//seq-lemale
/dev-stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Latmid BA; Site: 1; Not
I; Site-2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(GT) primer [5',
AAGTCAGAGAAATTCGCGCCGACGAGAAATTTTTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Latmid BA vector
library went through one round of normalization. Library
constructed by Bento Soares and M.Patima Bonaldo."
BASE COUNT      154 a      95 c      79 g      167 t      1 others

```


23

AUTHORS
Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A.,
Sano, M., Miyata, A. and Nishiyagi, A.

Page 8

Wed Sep 29 14:27:50 1999

us-09-030-606-224.rst

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 23:44:03 ; Search time 2910.9 Seconds
(without alignments)
349.617 Million cell updates/sec

Title: US-09-030-606-224
Perfect score: 320

Sequence: 1 CCCCTGAGGCTTCTGTTA.....TTTARACTCMGCATTGTGAC 320

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database

```

Genbml: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_om: *
4: gb_ov: *
5: gb_pat: *
6: gb_ph: *
7: gb_pl1: *
8: gb_pl2: *
9: gb_pr1: *
10: gb_pr2: *
11: gb_pr3: *
12: gb_ro: *
13: gb_st: *
14: gb_sts: *
15: gb_sy: *
16: gb_un: *
17: gb_v1: *
18: em_fun: *
19: em_hcg: *
20: em_hum1: *
21: em_hum2: *
22: em_in: *
23: em_om: *
24: em_or: *
25: em_ov: *
26: em_pat: *
27: em_ph: *
28: em_pl: *
29: em_ro: *
30: em_sts: *
31: em_sy: *
32: em_un: *
33: em_v1: *
34: gb_hcg1: *
35: gb_hcg2: *
36: gb_in1: *
37: gb_in2: *
38: em_ba1: *
39: em_ba2: *
40: em_hum3: *
41: em_hum4: *
42: gb_pr4: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	45	14.1	10094	2	067573	067573 Methanococcus

2	43.8	13.7	199917	34	AC006755	AC006755 Caenorhab
3	43.2	13.5	14553	37	AE001391	AE001391 Plasmodiu
4	40.8	12.7	194410	35	AC005140	AC005140 Plasmodiu
5	40.8	12.7	219300	35	AC0065279	AC0065279 Plasmodiu
6	40.4	12.6	4013	1	SMDNAMRPR	X75429 S. aureus p1
7	40.4	12.6	29777	36	CEFA6C3	Z66563 Caenorhabdi
8	40	12.5	192151	35	AC007282	AC007282 Homo sapi
9	39.8	12.4	78338	42	AF129078	AF129078 Homo sapi
10	39.6	12.4	91029	34	HS72E17	AL035353 Homo sapi
11	39.6	12.4	312766	34	PFMLAP3	AL035476 Plasmodiu
12	38.8	12.1	89603	8	AC007659	AC007659 Arabidops
13	38.8	12.1	85534	8	ATPAC02388	AC003388 Arabidops
14	38.8	12.1	69037	12	AC003949	AC003949 Mus muscu
15	38.4	12.0	138888	10	HS800E24	Z29758 Human DNA s
16	38.4	12.0	261570	34	PFMLAP2	AL035445 Plasmodiu
17	38.4	12.0	121336	42	AC006458	AC006458 Homo sapi
18	38.2	11.9	7218	5	166494	166494 Sequence 14
19	38.2	11.9	87104	8	ATP5C23	AL049500 Arabidops
20	38.2	11.9	324429	34	CEY67H2	AL023425 Caenorhab
21	38.2	11.9	23952	36	CEP28D10	275552 Caenorhabdi
22	38	11.9	10214	2	MC62057	U62057 Mycoplasma
23	38	11.9	126041	11	AC004904	AC004904 Homo sapi
24	38	11.9	190434	11	AC005771	AC005771 Homo sapi
25	37.8	11.8	104871	11	AC004584	AC004584 Homo sapi
26	37.8	11.8	116019	11	AC004629	AC004629 Homo sapi
27	37.8	11.8	551	36	DDCOXV	X64771 D. discoideu
28	37.8	11.8	81023	37	AC001659	AC001659 Drosophil
29	37.6	11.8	84440	7	ABD10075	ABD10075 Arabidops
30	37.6	11.8	118086	7	ATP6H11	AL021684 Arabidops
31	37.6	11.8	180000	35	AC004578	AC004578 Homo sapi
32	37.6	11.8	190165	35	AC006235	AC006235 Homo sapi
33	37.6	11.8	180117	35	AC007053	AC007053 Drosophil
34	37.6	11.8	165283	37	AC005711	AC005711 Drosophil
35	37.6	11.8	203199	37	AC005721	AC005721 Drosophil
36	37.4	11.7	4341	1	MHTYBPC	X62936 Mycoplasma
37	37.4	11.7	117381	11	AC006392	AC006392 Homo sapi
38	37.2	11.6	17056	3	MTDNCOMGN	Y11832 D. novemcinc
39	37.2	11.6	137785	9	HS345P10	282201 Human DNA s
40	37.2	11.6	32981	36	CE1KZ418	U00047 Caenorhabdi
41	37	11.6	11246	2	U67600	U67600 Methanococc
42	37	11.6	71934	9	HS159A15	AL022575 Homo sapi
43	36.8	11.5	220000	34	CEY111B2	Z29857 Caenorhabdi
44	36.8	11.5	110000	35	AC007126	AC007126 Homo sapi
45	36.8	11.5	147724	35	AC007273	AC007273 Homo sapi

ALIGNMENTS

RESULT	1
LOCUS	U67573
DEFINITION	U67573 10094 bp DNA BCT 20-MAY-1998
ACCESSION	U67573 L77117
NID	92826395
VERSION	067573.1
KEYWORDS	GI:28263395
SOURCE	
ORGANISM	Methanococcus jannaschii.
REFERENCE	Methanococcus jannaschii.
AUTHORS	Archaea: Euryarchaeota: Methanococcales: Methanococcaceae; Methanococcus jannaschii. 1 (bases 1 to 10094) Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D., Sutton,G.G., Blake,J.A., Fitzgerald,L.M., Clayton,R.A., Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J., Adams,M.D., Reich,C.L., Overbeek,R., Kirkness,E.F., Weinstock,K.G., Merriell,J.M., Glodek,A., Scott,J.D., Geoghagen,N.S., Weidman,J.F., Fuhrman,J.L., Nguyen,D.T., Utterback,T., Kelley,J.M., Peterson,J.M., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A., Roberts,K.M., Raine,B.B., Borodovsky,M., Klenk,H.P., Fraser,C.M., Smith,H.O., Woese,C.R. and Venter,J.C. Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
TITLE	

Query Match	13.5%	Score 43.2	DB 37	Length 14553
Best Local Similarity	57.7%	Pred. No. 0.91		
Matches 75	Conservative 1	Mismatches 54	Indels 0	Gaps 0
DB	38	ACCAATAGGACACACAAAAAGAAAGTTTGTACACTTTGATAGTAGGAGGTGTACCCC	97	
DB	6985	ACCAACAAAAAATAAATAATATACAAAGTTGAAAGATATTATTAAGAAAGCTTAAT	6926	
OY	98	TTACTCCCATCAAAAAAATGATATACGTGTTAAAGATATARAAGGACATATTTAT	157	
DB	6525	GTAAGAAGGATAAAAAATAATATATATATATATAAAGAAAAAAGCTAAATTTTAC	6866	
OY	158	CATATGTCT	167	
DB	6865	TTTTTTTTT	6856	
RESULT 4	AC005140	194410 bp	DNA	HTG 01-APR-1999
LOCUS	AC005140/c	Plasmodium falciparum chromosome 12, WORKING DRAFT SEQUENCE, 14		
DEFINITION	AC005140	unordered pieces.		
ACCESSION	AC005140	94558572		
NID	AC005140.3	GI:4558572		
VERSION	AC005140.3	HTG_PHASE1.		
KEYWORDS		malaria parasite P. falciparum.		
SOURCE		Plasmodium falciparum		
ORGANISM		Ekaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
REFERENCE		1 (bases 1 to 194410)		
AUTHORS		Hyman, R.W., Fung, E.L., Qin, F., Tamaki, T., Kuri, O.B. and Davis, R.W.		
JOURNAL		Plasmodium falciparum 3D7 chromosome 12		
REFERENCE		2 (bases 1 to 194410)		
AUTHORS		Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.		
TITLE		Direct Submission		
JOURNAL		Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology		
COMMENT		Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
		On Apr 2, 1999 this sequence version replaced gi:4337162.		
		* NOTE: This is a 'working draft' sequence. It currently		
		* consists of 14 contigs. The true order of the pieces		
		* is not known and their order in this sequence record is		
		* arbitrary. Gaps between the contigs are represented as		
		* runs of N, but the exact sizes of the gaps are unknown.		
		* This record will be updated with the finished sequence		
		* as soon as it is available and the accession number will		
		* be preserved.		
		1 13954: contig of 13954 bp in length		
		* 13955 14154: gap of unknown length		
		* 14155 32431: contig of 18277 bp in length		
		* 32432 32631: gap of unknown length		
		* 32632 105078: contig of 72447 bp in length		
		* 105079 105278: gap of unknown length		
		* 105279 117575: contig of 12297 bp in length		
		* 117576 117775: gap of unknown length		
		* 117776 130540: contig of 2565 bp in length		
		* 130541 130540: gap of unknown length		
		* 120341 132816: contig of 12276 bp in length		
		* 132817 133016: gap of unknown length		
		* 133017 135389: contig of 2373 bp in length		

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1	77526: contig of 77526 bp in length
* 77527	77726: gap of unknown length
* 77727	140331: contig of 62405 bp in length
* 140332	140331: gap of unknown length
* 140333	173508: contig of 33177 bp in length
* 173509	173708: gap of unknown length
* 173709	212797: contig of 39089 bp in length
* 212798	212997: gap of unknown length
* 212998	214904: contig of 1907 bp in length
* 214905	215104: gap of unknown length
* 215105	219200: contig of 4096 bp in length.

BASE COUNT	88210 a	22418 c	19532 g	88036 t	1004 others
ORIGIN					

Query Match	12.7%	Score 40.8;	DB 35;	Length 219200;
Best Local Similarity	47.2%	Pred. No. 2.4;		
Matches 117; Conservative	2;	Mismatches 129;	Indels 0;	Gaps 0;

QY 20 AGAAATAGTACAGTTACACCATAAGAACAAAAAGTTGTGCATTGTA 79
| | | | | | | | | | | | | | |
Db 85299 ATAAAAAAAAATAAAAAAAAATAAAAAAAAATAAAAAAAAATAAA 85240

80 GTAGGAGTGTGTACCCCTTACTCCCATCAAAAAAAAAATGGATACATGTTAAAGCAT 139

140 ARAAGGCAATATTTTATCATATGTTCTAAAGAGAGAGAAATACTTCTC 199

200 RAATGGAACCCCTTAAGGTCCTTGATCTACTGAAGACACAATGTGGCCGTCATCCT 259

[illegible]

Db 85059 TCCTAAA 85052

RESULT 6			
SADNAMUPR/c			
LOCUS	SADNAMUPR	4013 bp	DNA
			BCT
			01-AUG-1994

DEFINITION	<i>S. aureus</i> plasmid encoded DNA, mup K gene.
ACCESSION	X75439
NID	9438226
VERSION	CT.438226
VERSION	X75439.1

KEYWORDS	isoleucyl-tRNA synthetase.
SOURCE	<i>Staphylococcus aureus</i> .
ORGANISM	<i>Staphylococcus aureus</i>

REFERENCE
1 (bases 1 to 4013)
Hodges T F, Currock S B, Duha K C, Morris B, Eulimactor D B
Bacteria; Firmicutes; Bacillus/clostridium group; Bacillaceae;
Staphylococcus.

TITLE
and Gross, M.S.
Molecular characterization of the gene encoding high-level
mupirocin resistance in *Staphylococcus aureus* J2870

BOOKS
MEDLINE
94346838
REFERENCE
2 (bases 1 to 4013)
AUTHORS
Hodges T F

Direct Submission

true right end of clone F08B12 is at 11562 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence 268104. The end of this sequence (29718..29777) overlaps with the start of sequence 268107.

FEATURES

Source

location/Qualifiers

1..29777

/organism="Caenorhabditis elegans"

/chromosome="X"

/clone="F46C3"

complement(5728..7331)

gene

CDS

complement(join(5728..5791,5836..6003,6061..6189,6239..6419,6467..6608,7046..7192,7242..7331))

/gene="F46C3.2"

/codon_start=1

/protein_id="CA91468.1"

/db_xref="PID:e1346871"

/db_xref="GI:3877185"

/db_xref="GI:3877185"

/db_xref="SPTREMBL:O20457"

/translation="MMLTVVGQFQPLVTKLTKGKMLRMQLLTHVSDVLDVYTN

DKHOKRALNLREGVLTDFECFNTSTHVGPERQVPMPEVEMRLIOSEQDLQNE
KPMILKFTMKHVSASFNDHVRADLRSLTARVMPITWTLFESEMTLATASHI
WSQVNTSDELQETLSMTLPITLIVSSILIGMLIFVYVMGFLCCENNESMLVTK
DLVSSLYHCELENKRFDRIRYEVKGOEILVWTLMLAVAHSDINITYALVCVLI
NGFGLMFAHAEFRRCRRDGTSTIAR"

complement(9727..26596)

gene

CDS

/gene="hum-4"

complement(join(9727..9807,10085..10186,10257..10423,
10474..10616,11034..11173,11223..11383,11430..11604,
11647..11780,11846..11979,12137..12416,12469..12577,
12628..12776,12901..13082,13137..13261,13861..13979,
14025..14261,14311..14431,14665..14838,14887..15013,
15220..15302,15960..16008,16080..16201,16482..16676,
16726..17201,17383..17609,17720..17873,17922..18046,
18094..18288,18343..18477,18522..18682,18876..19118,
19162..19330,19990..20160,20230..20339,20385..20525,
20573..20737,20784..21543,21589..21783,21827..21933,
23751..24069,25817..25911,26033..26348,26397..26566))

/note="Similarity to myosin: CDNA EST yk24944.5 comes from
this gene; CDNA EST yk270h6.5 comes from this gene; CDNA
EST yk284c3.5 comes from this gene; CDNA EST yk353c4.5
comes from this gene; CDNA EST yk470b4.5 comes from this
gene; CDNA EST yk24944.3 comes from this gene; CDNA EST
yk284c3.3 comes from this gene; CDNA EST yk470d4.3 comes
from this gene"

/codon_start=1
/protein_id="CA91469.1"
/db_xref="PID:e1346870"
/db_xref="PID:g3877186"
/db_xref="GI:3877186"

/db_xref="SPTREMBL:O20456"

/translation="MSTEVALLNPDDHHTEDHIEANISRRKNSIACYNNMOTIPLNF

NFNSKPAIFKISVKGDLRLINDPEATAPRTNMTATNAPVSTFSPSCGELA
VYNINCESLKAROTDHOIPKKNVYTKICEVQRLGNSHGLQSPDPYUILLBIYAN
IHLLEEFPSSEETLFLLESTAEKFKKHVRKLPFCSSSEIYSLNMCIMDRWN
EGYVMDNFTRLMLSDTRWEVETNOFILRWGRMLDLEFELLQSDSTRIGLVS
EMESSTPIHGFEKTKSPSWLASLILKPKQTKASNVYFNDFNDQSEQLSV
ETSTSGVNAVAKNAKNIKFNMSNASSIVGSGSGSKSYNFAKRYLQSQRKY
STKSSSIEFVFSFGCAKTLKNDKATRFSGSIDLLYKRVNLTGLNKTYPLEPVY
ISOKPRGRNPIFYEYHGLSDMKAKFGIGLOKFTYINGNSSENQIDHYNEKVL
ESALHJGSDHDCMSITKITSLHIGNITFRKRNPNVQDVEIGNELMWLW
LLEVDQQLVKEFLPTSEDGSTELNALDRDSFAMVIEELKAVNLIRGLQKLS
LHGVSTILDHGFGEKYNNGVEEFLINSVNERLENFVKCFDQDLIDYKDCISD
YKVNSLENGKTVLEFKKPYGLSLILDECKPFGHEYTELECNHNGDSAYGA
RNKELEGVHRCIGTMYNTDFPARKRILISAVOLMNSKNPIIIGLFEESGN
TSDIIVSOAOPLVLAGADINDAKIVSHVHPRCKSNNSRSTEDPIYVNRQKNL
LAEILSRVKGTPYKISKITFARQYRCLDPEDICQCEKSTIIDYIQGGQKEDDE
KIGTEYVFLERLAERYDQNLKICGDAIILIKNMSFVAQKYKRRRAIILQSG
LRGKARADYILIKREMEKAIKGRMKRKRRLDAVHAQALGTNSGLOQLVGYIDINE

gene

CDS

complement(2750..3286),complement(2424..2700),
complement(2206..2376),complement(1899..2155),
complement(1748..1849),complement(1130..1425),
complement(1709..1079),complement(385..659),
complement(105..304),complement(268104.1:26738..27119),
complement(268104.1:26574..26660))

/gene="F46C3.1"

/note="Similar to serine/threonine kinase (2 domains):
CDNA EST EMBL:D34010 comes from this gene; CDNA EST
EMBL:D36999 comes from this gene; CDNA EST EMBL:D64541
comes from this gene; CDNA EST yk459d11.3 comes from this
gene; CDNA EST yk459d11.5 comes from this gene; CDNA EST
yk459d11.5 comes from this gene; CDNA EST yk22a45.5 comes
from this gene; CDNA EST yk251c3.3 comes from this gene;
CDNA EST yk427c7.5 comes from this gene"

/codon_start=1

/protein_id="CA91470.1"

/db_xref="PID:e1346871"

/db_xref="PID:g3877187"

/db_xref="GI:3877187"

/db_xref="SPTREMBL:O19192"

/translation="MSVYIVLVAGFLFLMALVFNAGQYIIDIEVSSCGQYVON
AGDDNNOVNIITATIDGVVTLGSGTGMIRVDEAPLRLKTLISGSDIDGSLT
QMPITDGRFSTTHNNLEPILPITDLSLESTIRLGOAVAGSKSYTRKGDLPFG
EOKTEGSCSGEDETPNPITILIRTNISRAMDTLGGIEMWNLSTAEVTLVG
GITSPTLVSOVKLLQPPDIVAVADVYNEEKTVNDGILVWVOYGGQIGETISL
DPSNIETQVQREOHNLQOTSLMYTSGFPGIIOISPAKNNLQKRMALNEL
SMTLELNPFCANETRELAIVNDEDTLVLHNAFRSQKALDEKLSGSSAR
KLIILASDTEVAGIETENLSTVYSKSDVYLVLESPOKRVFPIITLMORI
FSTINPTAVSTLACIGTVAVVYNNKIAKSSPRMIEHLSSTESATBSAIRTPT
SFAPIDELIREVBSDSLTPIDGAIHRPMPLEISNLETHQPIRKVQVQKVD
IDIDEDSFNSDEKRLRLNRTISRSLEGTISRFANEFEVKVIGHGGFVPRAOI

Query Match 12.6%; Score 40.4; DB 36; Length 29777;
 Best Local Similarity 50.0%; Pred. No. 3.5;
 Matches 95; Conservative 2; Mismatches 99; Indels 0; Gaps 0;

OY 43 ATAGAACACAAAAAGAAAGTTTGTGACATTTAGTAGGAGTGTACCCCTTACT 102
 DB 1473 ACACAAAAAATAGATGACATTTGTGAAATTTAGTTAATTAATAGCAAAACA 1532
 OY 103 CCCCATCAAAAAAATGATGATGATTAAGATTAAGAGGCAATTTATCATAT 162
 DB 1533 GTCTATTCAAAAATGCAACGCTTGAAAGAAAGGGAATGCTTTTGGTTTGTGA 1592
 OY 163 GTTCTAAAGAGAGAGAGAAATTAATCTTCTTAAATGAGCCCTTAAGAGTGC 222
 DB 1593 CGTATAGATCCGAATGCTCTTAATAATATTTCAAAATGGAATCTTAATAATGC 1652
 OY 223 TTGATATCTG 232
 DB 1653 TTGAATATG 1662

RESULT 8
 AC007282 AC007282 192151 bp DNA 05-JUN-1999
 LOCUS Homo sapiens clone NH0484N09, WORKING DRAFT SEQUENCE, 4 unordered
 DEFINITION pieces.

ACCESSION AC007282
 NID 95001487
 VERSION AC007282.2 GI:5001487
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eutheria; Primates; Catarrhini; Homiidae; Homo.
 AUTHORS 1 (bases 1 to 192151)
 TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 192151)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (09-APR-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

COMMENT On Jun 5, 1999 this sequence version replaced gi:4580496.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES
 source
 1. 192151
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NH0484N09"

BASE COUNT 62623 a 38154 c 35320 g 56003 t 51 others

Query Match 12.5%; Score 40; DB 35; Length 192151;
 Best Local Similarity 50.0%; Pred. No. 3.6;
 Matches 94; Conservative 2; Mismatches 92; Indels 0; Gaps 0;

OY 29 TACAGTACACCAATAGAACACAAAAAGAAAGTTTGTGACATTTAGTAGGAGT 88
 DB 75422 TAGAGTAAGACCTCTCTCCAAAAAAGAGCTATGAGAGCAGTTAATATA 75481
 OY 89 GTGATCCCTTACTCCCATCAAAAAAATGATGATGTTAAGATTAAGAGGCA 148
 DB 75482 TTTAAGTCCAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 75541
 OY 149 AATTTTATCATATGTTCTTAAAGAGAGAGAAATTAATCTTCTCTAAATGGA 208
 DB 75542 AACAACTAATAAATAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 75601
 OY 209 GCCCTTAA 216
 DB 75602 TTAATTAA 75609

RESULT 9
 AF129078/c AF129078 78338 bp DNA 04-MAR-1999
 LOCUS Homo sapiens chromosome 21q22.1 PAC B22238, complete sequence.
 DEFINITION AF129078
 ACCESSION AF129078
 NID 94337089
 VERSION AF129078.1 GI:4337089
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 AUTHORS 1 (bases 1 to 78338)
 TITLE Direct Submission
 JOURNAL Submitted (17-FEB-1999) Genome Analysis, Institute of Molecular
 Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
 LOCATION/Qualifiers

FEATURES
 source

1. 78338

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"

/map="21q22.1"
 /clone="PAC B22238"

/complement(1..153)

/rpt_family="MT1A1"

/evidence-not_experimental

/complement(332..407)

/note="MZF, score = 83%"

/evidence-not_experimental

1315..1687

/rpt_family="THEIC"

/evidence-not_experimental

1817..1932

/note="GRAIL, score = 59.000%, comment = good"

/evidence-not_experimental

1850..1928

/rpt_family="MT1A1"

/evidence-not_experimental

1860..1914

/note="Xpound exon prediction, score = 65% (0%)"

/evidence-not_experimental

complement(1933..2079)

/rpt_family="MT1A1"

/evidence-not_experimental

complement(2080..2341)

/rpt_family="AluJo"

/evidence-not_experimental

complement(2342..2549)

/rpt_family="MT1A1"

/evidence-not_experimental

2721..3167

/rpt_family="LRR1"

/evidence-not_experimental

exon complement(2762, .2844)
 /note="MZF, score = 60.4%"
 /evidence-not_experimental
 repeat_region 3260, .3470
 /rpt_family="LFR28"
 /evidence-not_experimental
 complement(3471, .3538)
 /rpt_family="MLT1A2"
 /evidence-not_experimental
 repeat_region 4631, .4700
 /note="homology = 70.00%, score = 21, counts = 10"
 /rpt_type="tandem"
 /rpt_unit="aaagaa"
 /evidence-not_experimental
 complement(4825, .4872)
 /note="GRAIL, score = 84.000%, comment = excellent"
 /evidence-not_experimental
 exon 5249, .5414
 /note="Genscan, score = 4.31%, comment = Terminal_exon 166
 bp frame: 2 phase: 1"
 /evidence-not_experimental
 repeat_region 6153, .6187
 /rpt_family="L2"
 /evidence-not_experimental
 complement(6600, .6701)
 /rpt_family="L2"
 /evidence-not_experimental
 repeat_region 6723, .7190
 /rpt_family="MLT1H"
 /evidence-not_experimental
 complement(6941, .6992)
 /note="Xpound exon prediction, score = 66% (0%)"
 /evidence-not_experimental
 repeat_region 7367, .7665
 /rpt_family="L2"
 /evidence-not_experimental
 complement(9762, .9962)
 /rpt_family="MER3"
 /evidence-not_experimental
 repeat_region 9664, .10150
 /rpt_family="MIR"
 /evidence-not_experimental
 complement(10362, .10639)
 /rpt_family="Alusg"
 /evidence-not_experimental
 exon complement(10363, .10619)
 /note="MZF, score = 81.3%"
 /evidence-not_experimental
 complement(10895, .11136)
 /rpt_family="MIR"
 /evidence-not_experimental
 repeat_region 13416, .13515
 /rpt_family="MER5A"
 /evidence-not_experimental
 repeat_region 14240, .14314
 /rpt_family="MER5A"
 /evidence-not_experimental
 exon 14515, .14671
 /note="GRAIL, score = 95.000%, comment = excellent"
 /evidence-not_experimental
 exon 14515, .14671
 /note="MZF, score = 56.7%"
 /evidence-not_experimental
 repeat_region 14766, .14791
 /note="111 with 96% homology to 112"
 /rpt_type="inverted"
 /evidence-not_experimental
 complement(14808, .15551)
 /rpt_family="L1PA2"
 /evidence-not_experimental
 repeat_region complement(14957, .15105)
 /note="GRAIL, score = 63.000%, comment = good shadow"
 /evidence-not_experimental

exon complement(15431, .15537)
 /note="GRAIL, score = 47.000%, comment = marginal"
 /evidence-not_experimental
 repeat_region 15552, .15577
 /note="112 with 96% homology to 111"
 /rpt_type="inverted"
 /evidence-not_experimental
 repeat_region 16981, .17278
 /rpt_family="Alusg"
 /evidence-not_experimental
 complement(17569, .18083)
 /rpt_family="MER51B"
 /evidence-not_experimental
 complement(18364, .18887)
 /rpt_family="L1ME2"
 /evidence-not_experimental
 complement(18884, .19040)
 /rpt_family="L1ME2"
 /evidence-not_experimental
 repeat_region complement(19041, .19170)
 /rpt_family="F1M.C"
 /evidence-not_experimental
 exon 19058, .19149
 /note="MZF, score = 75.9%"
 /evidence-not_experimental
 complement(19171, .19315)
 /rpt_family="L1ME2"
 /evidence-not_experimental
 complement(19314, .19857)
 /rpt_family="L1MEC"
 /evidence-not_experimental
 complement(19837, .19944)
 /rpt_family="L1MEC"
 /evidence-not_experimental
 repeat_region complement(19986, .20285)
 /rpt_family="Alusx"
 /evidence-not_experimental
 exon 20834, .20904
 /note="MZF, score = 52.3%"
 /evidence-not_experimental
 exon 20834, .20904
 /note="GRAIL, score = 70.000%, comment = good shadow"
 /evidence-not_experimental
 repeat_region complement(21748, .23646)
 /rpt_family="L1PA2"
 /evidence-not_experimental
 complement(21906, .22046)
 /note="GRAIL, score = 69.000%, comment = good shadow"
 /evidence-not_experimental
 repeat_region 23646, .23832
 /rpt_family="L1P"
 /evidence-not_experimental
 repeat_region 23833, .24258
 /rpt_family="L1R7"
 /evidence-not_experimental
 repeat_region 24267, .24324
 /rpt_family="L1P"
 /evidence-not_experimental
 exon 24325, .24762
 /rpt_family="L1P"
 /evidence-not_experimental
 complement(24341, .24410)
 /note="MZF, score = 87.1%"
 /evidence-not_experimental
 repeat_region 24763, .25071
 /rpt_family="AluJo"
 /evidence-not_experimental
 repeat_region 25072, .25617
 /rpt_family="L1P"
 /evidence-not_experimental
 exon 25096, .25247
 /note="GRAIL, score = 71.000%, comment = good"
 /evidence-not_experimental

RESULT 12
AC007659/c
LOCUS
DEFINITION Arabidopsis thaliana chromosome II BAC T14P1 genomic sequence,
complete sequence.
ACCESSION
AC007659
NID
VERSION AC007659.1 GI:4895233
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsi.
1 (bases 1 to 89603)
REFERENCE
AUTHORS
Goodman,H.M., Agyare,F., Belmonte,S., Dolan,M., Kim,U. and
Morris,J.W.
TITLE
Arabidopsis thaliana chromosome II BAC T14P1 genomic sequence
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 89603)
AUTHORS
Lin,X.
TITLE
Direct Submission
JOURNAL
Submitted (27-MAY-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
COMMENT
Address all correspondence to:
Howard Goodman
Mass. General Hospital
Boston MA 02114
goodman@mh.harvard.edu

This BAC was annotated in collaboration with TIGR group.
Contact:
Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr.
Rockville, MD 20850, USA
e-mail: xlin@tigr.org

BAC clone T14P1 is from Arabidopsis chromosome II and is near the
molecular marker m336.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of three methods: Gene
prediction programs including GRATL (available by anonymous ftp
from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
Washington), Genscan (Chris Burge,
http://www.cbs.dtu.dk/netpene/cbsnetpene.html), and NetPlantGene
(http://www.cbs.dtu.dk/netpene/cbsnetpene.html), searches of the
complete sequence against a peptide database and the Arabidopsis
EST database at TIGR (http://www.tigr.org/cdb/at/est.html).
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by more
than two gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
genomic sequence that are not annotated as genes but have predicted
exons by GRATL are annotated as misc features.
Location/Qualifiers
1..89603
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"
/map="m336"
/clone="T14P1"
complement(1..539)
/gene="T14P1.1"

misc_feature
1..1868
/note="Overlap with BAC clone F4L23
(AC002387:121061..122928)."
gene
complement(1..539)
/gene="T14P1.1"
/note="F4L23.31; identical to GB:AA82643"
complement(97..501)
/gene="T14P1.1"
/note="unknown protein"
CDS
/codon_start=1
/protein_id="AA032819.1"
/db_xref="PID:g4895234"
/db_xref="GI:4895234"
/translation="MASKALAVALLTLTLNLEFPTVTSRCPPTPKPKRPSPK
ARAKPCPDPTLKLGCADLGLVNVVGSPPRTCCITLGLIANEAIVCTALK
ANVGLNVPIIDLTLNLCGRVPHGFC"
733..842
misc_feature
/note="exon predicted by xgrall, quality
marginal shadowexon"
misc_feature
complement(791..852)
/note="exon predicted by xgrall, quality marginal"
misc_feature
complement(882..896)
/note="exon predicted by xgrall, quality marginal"
complement(join(1180..1571,1671..1723,2116..2171,
2294..2341,2499..2594))
/gene="T14P1.2"
complement(1180..2594)
/gene="T14P1.2"
/note="F4L23.32; identical to GB:AA82646. Two ESTs do not
appear to be completely spliced"
complement(join(1352..1571,1671..1723,2116..2171,
2294..2333))
/gene="T14P1.2"
/codon_start=1
/product="putative symbiosis-related protein"
/protein_id="AA032820.1"
/db_xref="PID:g4895235"
/db_xref="GI:4895235"
/translation="MNRGSLFKKDDDEKKRAENGRIREYPPRIPIVKAKESEY
NIDKKLYLPSDGLVGVYVIRRRRLSAEKALFIVDNLPTGLMSVYEKKD
EDGLYLYTSGENFGASIS"
complement(3021..3042)
/rpt_family="TA_rich"
complement(3443..3571)
/note="exon predicted by xgrall, quality excellent"
3451..3563
/note="exon predicted by xgrall, quality
marginal shadowexon"
misc_feature
complement(4034..4074)
/note="exon predicted by xgrall, quality good"
misc_feature
complement(4109..4140)
/note="exon predicted by xgrall, quality marginal"
4194..4302
/rpt_family="(TAA)n"
complement(4657..4723)
/note="exon predicted by xgrall, quality good"
complement(5290..5327)
/rpt_family="(TA)n"
complement(5891..5921)
/note="exon predicted by xgrall, quality
marginal shadowexon"
6344..6406
/note="exon predicted by xgrall, quality marginal"
complement(6361..6412)
/note="exon predicted by xgrall, quality
marginal shadowexon"
6589..6645
/rpt_family="(GAAA)n"
complement(6722..6814)
/rpt_family="(GAAA)n"
6845..8942
/gene="T14P1.3"
complement(1..539)
/gene="T14P1.1"
6845..8942
gene

```

CDS
    /gene="t14p1.3"
    /gene=".8767
    /gene="t14p1.3"
    /note="unknown protein"
    /codon_start=1
    /protein_id="AAD32821.1"
    /db_xref="pid:q4895236"
    /db_xref="gi:4895236"
    /translation="MPLSEFRFGEGVEFLSSSSFFSDQKTSNDKTEAKOEDLGY
    VYGRPEPTSVLDADKSPSLASTSTTTITLSSSGGGGTTVNTVTAGDDNNNK
    CSOMGDDLDLVLSASSPGQEQSLRLMDPGSACVFPDGFSGFSGSPVAPADPG
    SNLCNDFPEQETNPAPALINPSNCLFNPNPLSPARKFNSSGSLHQPFPLSDPDP
    HDVROHOFQEPFHYNNQOQFSSSSSTAVAMPVPSPGAGDQOQSVITIQLEFNA
    ELIGTGNNNGDHVTAGGILARLNHNTSSNKKSPFORASHIAEALSLIHNSA
    PLITPENILRTAAVRSFSETSPLOFQFNTAKSLSSCSMGSGFDRTHIIDPNVGY
    GGOMSLMOELASGVGGRRRNASSKLTVRPPSTVDEFEETENTLTFACEV
    KLFETELLSVELLNPATWPLSLKSEKEALAVNPVNSVASGILPLILRELKLP
    NIVVOSDRGDRNDAPFNVAVHSLQYHSLSELDANODSSIERHWVQPSLEKL
    LMKRHRMERSPPWRILFQCGFSPASLSQMAEQAECLLQRPVGRFHEKRSSTLY
    MCMORRELVTVSAMKC"
    10014..10037
repeat_region
    /rpt_family="AT-rich"
    /note="exon predicted by xgrail, quality good"
    complement(10292..10333)
    /note="exon predicted by xgrail, quality
    marginal.shadowexon"
    10433..10467
repeat_region
    /rpt_family="AT-rich"
    10954..11000
    /rpt_family="(CAAAA)n"
    complement(join(<11530..11621,11704..11819,11904..12037,
    12119..12243,12366..12556,12650..12939,13131..13475))
    /gene="t14p1.4"
    complement(<11530..>13475)
    /gene="t14p1.4"
    complement(join(11530..11621,11704..11819,11904..12037,
    12119..12243,12366..12556,12650..12939,13131..13475))
    /gene="t14p1.4"
    /codon_start=1
    /product="putative phosphatidate cytidyllyltransferase"
    /protein_id="AAD32822.1"
    /db_xref="pid:q4895237"
    /db_xref="gi:4895237"
    /translation="MTNTTLPLRHHDDSEILLFOENSVFPLFVFAKGTNMAFEA
    ELYLSRCTCRSTCRSTKRPLIRPPLISGLRREDTLPRLRSGLGRISVRRLITAVAR
    AASDQJLDDHSGIDRIHNLQVEDKOKASOLKRVIFGIGLPGCVLAGGWA
    FTVALLASVFIQSREYFELVRSKIGAKGMPPRVRSVCSVICALMPLITLYFGND
    IIVTSAFVVAIALIVQSGSPRAQSLSTMGFLYCGYLPSEFWKRLCGLAAPALNTG
    IGRTPWILGGGAHMTVGAVATLISFSGVIATDFALGKTEGRTPLTISIPKRTWE
    GTVGVGCIATITLISKYLSPQSLFSSVAFGLNPGSVGCDLLESMIKKADAGKD
    SSSLIPRHGIGIIDRVDSYFTGALATSEFTKSLKYG"
    complement(join(113650..14019,14258..14335,14868..14918,
    14999..15157,15239..15354,15460..15541,15621..15691,
    15800..15877))
    /gene="t14p1.5"
mRNA
Query Match      12.13; Score 38.8; DB 8; Length 89603;
Best Local Similarity 48.58; Pred. No. 7.2;
Matches 100; Conservative 2; Mismatches 104; Indels 0; Gaps 0;

```

```

OY      213  TTAAGTGTCTTGATCACTGAAGAC 238
DB      84425  TGCATATTCATTTCTTCATGAC 84400

RESULT 13
ATAC002388/c
LOCUS   Arabidopsis thaliana chromosome II BAC T13E15 genomic sequence,
DEFINITION complete sequence.
ACCESSION AC002388
NID       93420042
VERSION   AC002388.1
KEYWORDS  GI:3420042
SOURCE    HTG.
ORGANISM  thale cress.
           Arabidopsis thaliana
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
           eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
           Arabidopsids.
           1 (bases 1 to 85534)
           Rounsley,S.D., Lin,X., Ketchum,K.A., Crosby,M.L., Brandon,R.C.,
           Sykes,S.M., Mason,T.M., Kerlavage,A.R., Adams,M.D., Somerville,C.R.
           and Venter,J.C.
           Arabidopsis thaliana chromosome II BAC T13E15 genomic sequence
           Unpublished
           2 (bases 1 to 85534)
           Rounsley,S.D. and Lin,X.
           Direct Submission
           Submitted (25-JUL-1997) The Institute for Genomic Research, 9712
           Medical Center Dr., Rockville, MD 20850, USA, Email:
           rounsley@tigr.org
           3 (bases 1 to 85534)
           Rounsley,S.D.
           Direct Submission
           Submitted (27-AUG-1997) The Institute for Genomic Research, 9712
           Medical Center Dr., Rockville, MD 20850, USA
           4 (bases 1 to 85534)
           Rounsley,S.D.
           Direct Submission
           Submitted (15-AUG-1998) The Institute for Genomic Research, 9712
           Medical Center Dr., Rockville, MD 20850, USA
           On Aug 15, 1998 this sequence version replaced gi:2344885.
           Address all correspondence to:
           Steve Rounsley
           The Institute for Genomic Research
           9712 Medical Center Dr,
           Rockville, MD 20850,
           USA
           e-mail: rounsley@tigr.org
           BAC clone T13E15 is from Arabidopsis chromosome II and contains the
           molecular marker m336.
           The orientation of the sequence is from SP6 to T7 end of the BAC
           clone.
           Genes were identified by a combination of three methods: Gene
           prediction programs including GRAFT (available by anonymous ftp
           from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
           Washington), Genscan (Chris Burge,
           http://www.cds.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
           (http://www.cds.dtu.dk/netgene/cdsnetgene.html), searches of the
           complete sequence against a peptide database and the Arabidopsis
           EST database at TIGR (http://www.tigr.org/tdb/at/t.html).
           Annotated genes are named to indicate the level of evidence for
           their annotation. Genes with similarity to other proteins are named
           as isoforms. Genes without significant peptide similarity but with
           EST similarity are named as 'unknown' proteins. Genes without
           protein or EST similarity, that are predicted by more than two gene
           prediction programs over most of their length are annotated as
           'hypothetical' proteins. Genes encoding tRNAs are predicted by
           tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
           Simple repeats are identified by RepeatMasker (Arian Smith,
           http://ftp.genome.washington.edu/SW/RepeatMasker.html). Regions of
           genomic sequence that are not annotated as genes but have predicted

```


QY 153 TTTATCATATGTTCTTAAAGAGAGAGAGAAATACTACTTTCRAAATGAGAGCC 212
 DB 11068 TGTACTCTTCTCTTCACATTAATCCAAATGCTTCTCAAAACTTGATT 11009
 QY 213 TTAAGGTGCTTGATGACTGAGAGC 238
 DB 11008 TGCATATCCAAATTAATCTTCATGAC 10983

RESULT 14
 AC003949
 LOCUS AC003949 69037 bp DNA ROD 03-JUL-1998
 DEFINITION Mus musculus chromosome 19, clone D19-96.B7, complete sequence.
 ACCESSION AC003949
 NID 93287717
 VERSION AC003949.1 GI:3287717
 KEYWORDS HTG.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Euteleia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 69037)
 Birren, B., Fasmann, K., Linton, L., Nusbaum, C. and Lander, E.
 Mus musculus chromosome 19, clone D19-96.B7
 Unpublished

REFERENCE 2 (bases 1 to 69037)
 Birren, B., Fasmann, K., McKernan, K., Munro, C., Nusbaum, C.,
 Richardson, P., Lander, E., Allen, N., Baldwin, J., Barna, N.,
 Beckery, R., Cantu, C., Castle, A., Cooke, P., Daly, M., Depayre, E.,
 Devon, K., Dewar, K., Durette, B., Ferreira, P., Forrest, C., Gage, D.,
 Gardyna, S., Gensheimer, S., Geraltier, K., Gilmartin, T., Gray, D.,
 Hagos, B., Halphen, I., Harris, K., Horton, L., Howland, J., Huang, J.,
 Hul, L., Jacotot, L., Linton, L., Mackenzie, J., Marquis, N.,
 McDermett, J., McEwan, P., McGurk, A., Meldrum, J., Molla, M.,
 Morris, W., Morrow, J., Nachman, A., Naylor, J., O'Connor, T.,
 Pavlin, B., Peterson, K., Riley, R., Roberts, D., Rollins, G., Roy, A.,
 Sarnaik, A., Shlu, P., Shyam, R., Stillwell, J., Stone, C.,
 Strickland, C., Sydney, K., Tang, L., Vassiliev, H., Vo, A., Wu, Y.,
 Ye, W. J., Zemtseva, I., Zhao, J. and Zody, M.

TITLE Direct Submission
 JOURNAL Submitted (23-DEC-1997) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 69037)
 Birren, B., Fasmann, K., McKernan, K., Nusbaum, C., Richardson, P.,
 Lander, E., Allen, N., Baker, J., Baldwin, J., Barna, N., Beckery, R.,
 Boutwell, C., Byrne, S., Cantu, C., Castle, A., Cooke, P., Daly, M., J.,
 Depayre, E., Devon, K., Dewar, K., Durette, B., Etemadi, S.,
 Ferreira, P., Forrest, C., Gage, D., Gardyna, S., Gensheimer, S.,
 Geraltier, K., Gilmartin, T., Gray, D., Hagos, B., Harris, K.,
 Horton, L., Howland, J., Hul, L., Jacotot, L., Linton, L.,
 Mackenzie, J., Marquis, N., McEwan, P., McGurk, A., Meldrum, J.,
 Molla, M., Morris, W., Morrow, J., Nachman, A., Naylor, J., O'Connor, T.,
 Pavlin, B., Peterson, K., Ranganath, S., Riley, R., Roberts, D.,
 Rollins, G., Rossello, R., Roy, A., Shyam, R., Sooboo, S., Stillwell, J.,
 Stone, C., Strickland, C., Sydney, K., Tang, L., Vassiliev, H., Vo, A.,
 Wagner, A., Wheeler, J., Wu, Y., Ye, W. J., Zemtseva, I., Zhao, J. and
 Zody, M.

TITLE Direct Submission
 JOURNAL Submitted (10-FEB-1998) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 4 (bases 1 to 69037)
 Birren, B., Fasmann, K., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
 Baker, J., Baldwin, J., Barna, N., Beckery, R., Benn, J., Boatman, C.,
 Boutwell, C., Brown, A., Castle, A., Cerny, J., Cooke, P., Depayre, E.,
 Devon, K., Dewar, K., Donegan, L., Etemadi, S., Ferreira, P.,
 Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Gardyna, S.,
 Gensheimer, S., Geraltier, K., Gilmartin, T., Grant, G., Hagos, B.,
 Harris, K., Horton, L., Howland, J., Hul, L., Jacotot, L., Kann, L.,
 MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
 Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
 Nachman, A., Naylor, J., O'Connor, T., Pavlin, B.,
 Peterson, K., Riley, R., Roberts, D., Rossello, R., Roy, A., Shyam, R.,

TITLE
 JOURNAL
 COMMENT
 FEATURES
 source

Stange-Thomann, N., Stillwell, J., Stojanovic, N., Stone, C.,
 Strickland, C., Subramanian, A., Tormella-Willer, I., Vassiliev, H.,
 Vo, A., Wagner, A., Wang, B., Wheeler, J., Wu, Y., Ye, W. J., Zhao, J. and
 Zody, M.

Direct Submission
 Submitted (03-JUL-1998) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 3, 1998 this sequence version replaced g1.2865213.
 All repeats were identified using RepeatMasker: Smit, A.F.A. &
 Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Location/Qualifiers
 1..69037
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="D19-96.B7"
 /clone_11b="unknown"
 /map="19"
 /chromosome="19"
 1..1432
 /rpt_family="T1_MM"
 complement(1519..1588)
 /rpt_family="AT_rich"
 complement(1870..1912)
 /rpt_family="AT_rich"
 1976..2079
 /rpt_family="TACGA)n"
 2296..2327
 /rpt_family="AT_rich"
 complement(2450..2749)
 /rpt_family="T1MA6"
 3166..3256
 /rpt_family="MTE"
 complement(3266..3331)
 /rpt_family="AT_rich"
 complement(3466..3516)
 /rpt_family="AT_rich"
 complement(4377..4399)
 /rpt_family="(GA)n"
 complement(4400..4459)
 /rpt_family="(TA)n"
 4495..4730
 /rpt_family="T1"
 4780..4825
 /rpt_family="(TAAA)n"
 4834..4896
 /rpt_family="(CA)n"
 5070..5461
 /rpt_family="T1"
 complement(5472..8049)
 /rpt_family="T1"
 complement(8255..8395)
 /rpt_family="RTR10"
 complement(9896..13758)
 /rpt_family="TAP"
 complement(15181..15321)
 /rpt_family="RTR10"
 complement(15322..16714)
 /rpt_family="T1"
 17080..17640
 /rpt_family="Txs"
 complement(17648..17681)
 /rpt_family="(TA)n"
 complement(17683..17804)
 /rpt_family="(TACGA)n"
 17814..19919
 /rpt_family="Txs"
 19644
 /note="Probably C."
 complement(21117..21225)
 /rpt_family="AT_rich"
 21286..21432
 /rpt_family="T1_MM"

```

repeat_region      21434. .21497
                    /rpt_family="CAAAA)n"
repeat_region      21719. .21743
                    /rpt_family="TAAA)n"
                    complement(21747..21869)
repeat_region      /rpt_family="AT_r1ch"
                    complement(22138..22846)
repeat_region      /rpt_family="Lx4"
                    complement(22875..23251)
repeat_region      /rpt_family="L1_MM"
                    complement(23235..23520)
repeat_region      /rpt_family="Lx4"
                    complement(23520..23698)
repeat_region      /rpt_family="L1"
                    complement(23926..24056)
repeat_region      /rpt_family="Lx3"
                    complement(24353..24390)
repeat_region      /rpt_family="AT_r1ch"
                    complement(25276..25307)
repeat_region      /rpt_family="Lx4"
                    complement(25416..26656)
repeat_region      /rpt_family="L1_MM"
                    complement(2657..28391)
repeat_region      /rpt_family="Lx4"
                    complement(28448..28491)
repeat_region      /rpt_family="AT_r1ch"
                    complement(28945..29032)
repeat_region      /rpt_family="AT_r1ch"
                    complement(29898..29954)
repeat_region      /rpt_family="TAAA)n"
                    complement(29957..30863)
repeat_region      /rpt_family="Lx2B"
                    complement(30905..32970)
repeat_region      /rpt_family="L1"
                    complement(33018..33876)
repeat_region      /rpt_family="L1"
                    complement(33895..33998)
repeat_region      /rpt_family="L1"
                    complement(34617..34657)
repeat_region      /rpt_family="AT_r1ch"
                    complement(35139..35012)
repeat_region      /rpt_family="B1_MM"
                    complement(35253..36201)
repeat_region      /rpt_family="L1"
                    complement(36211..37107)
repeat_region      /rpt_family="Lx2B"
                    complement(37094..37414)
repeat_region      /rpt_family="Lx3"
                    complement(37417..37598)
repeat_region      /rpt_family="Lx2"
                    complement(37613..38174)
repeat_region      /rpt_family="L1V2"
                    complement(38118..38381)
repeat_region      /rpt_family="L1V2"
                    complement(38410..38945)
repeat_region      /rpt_family="Lx7"
                    complement(39080..39151)
repeat_region      /rpt_family="AT_r1ch"
                    complement(39249..39316)
repeat_region      /rpt_family="AT_r1ch"
                    complement(39401..39940)
repeat_region      /rpt_family="RMR6A"
                    complement(41499..41933)
repeat_region      /rpt_family="L1_MM"

```

```

Query Match      12.1% Score 38.8; DB 12; Length 69037;
Best Local Similarity 48.5% Pred. No. 7.4;
Matches 100; Conservative 2; Mismatches 104; Indels 0; Gaps 0;
QY      11 CTTCCTTACAAATAGTACAGTACCAATAGAACAAACAAAGAAAGTTGT 70
      ||||| ||| || ||| ||||| ||| || ||| ||| |||

```

```

Db      65351 CTTCCTTATTAATTAACACATTTTAACATGATAGAAAAATCTAGCAACATTGTG 65410
QY      71 GACATGTAGAGGAGGTGCTACCCCTTACTCCCATCAAAAAAATGATACATGG 130
      ||||| ||| ||||| ||||| ||||| |||||
Db      65411 TATGAAGAGTCAACATTAATAATTCCTAGGAGCATAAAAAATATTAATTAAGCATGT 65470
QY      131 TTAAGGATTAAGGCAATTTATCATATGTCTTAAGAGAGCAAGCAATAATAC 190
      ||||| ||| ||| ||| ||| ||| ||| |||
Db      65471 ATAAACATTAAGACAGCAAAATTTTACTCATTTCTTAATAAGATGACTAATGTCC 65530
QY      191 TACTTCTCRAATGAGACCCCTTAA 216
      ||||| ||| ||| ||| ||| ||| ||| |||
Db      65531 TATGCTCTTAACATTAAGCAATTA 65556

RESULT 15
HS800F24/c      13888 bp      DNA      PRI      12-MAY-1999
LOCUS      Human DNA sequence from clone 800F24 on chromosome 1q24, complete
DEFINITION      sequence.
ACCESSION      289758
NID      94775628
VERSION      299758.7 GI:4775628
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 138888)
REFERENCE      Submitted (12-MAY-1999) Sanger Centre, Hinxton, Cambridgeshire,
AUTHORS      Grafham, D.
TITLE      Direct Submmission
JOURNAL      CS10 ISA. UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonequest@sanger.ac.uk
On May 11, 1999 this sequence version replaced gi:4678515.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EM, EMBL; SW, SWISSPROT; Tr, TREMBL
IMPORTANT: This sequence is not the entire insert of clone 800F24.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone d11018K9 is at 138789 in this sequence.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key
This sequence was generated from part of bacterial clone contigs of
human chromosome 1, constructed by the Sanger Centre Chromosome 1
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
800F24 is from the library RPC14 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/VECTOR: PCRAC2.
FEATURES
source
1. 138888
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="800F24"
/map="q24"
/clone_lib="RPC14"
1. 1964
/note="L1 repeat: matches 3306. .5335 of consensus"
repeat_region      1967. .2267

```

repeat_region	/note="Alusc repeat: matches 1. .300 of consensus"	repeat_region	27202. .27345
repeat_region	/note="LIP2 repeat: matches 5413. .5484 of consensus"	repeat_region	/note="L2 repeat: matches 2559. .2710 of consensus"
repeat_region	2340. .3242	repeat_region	27849. .28052
repeat_region	/note="LIM1 repeat: matches 799. .1710 of consensus"	repeat_region	/note="MIR repeat: matches 1. .205 of consensus"
repeat_region	3551. .3836	repeat_region	28914. .29215
repeat_region	/note="LIM1 repeat: matches 491. .794 of consensus"	repeat_region	/note="Alusc repeat: matches 1. .310 of consensus"
repeat_region	4306. .4426	repeat_region	29892. .30582
repeat_region	/note="MIR repeat: matches 87. .216 of consensus"	repeat_region	/note="LIP16 repeat: matches 5476. .6155 of consensus"
repeat_region	5580. .5677	repeat_region	30596. .31979
repeat_region	/note="MIR repeat: matches 49. .146 of consensus"	repeat_region	/note="LIP16 repeat: matches 4744. .6155 of consensus"
repeat_region	7200. .7639	repeat_region	32709. .32764
repeat_region	/note="LIR16B repeat: matches 1. .464 of consensus"	repeat_region	/note="L2 repeat: matches 2654. .2709 of consensus"
repeat_region	8511. .8829	repeat_region	34717. .35145
repeat_region	/note="Alusc repeat: matches 1. .309 of consensus"	repeat_region	/note="L2 repeat: matches 1906. .2371 of consensus"
repeat_region	10092. .10203	repeat_region	35142. .35391
repeat_region	/note="L2 repeat: matches 2613. .2710 of consensus"	repeat_region	/note="L2 repeat: matches 2141. .2359 of consensus"
repeat_region	11207. .11257	repeat_region	35392. .35691
repeat_region	/note="MIR repeat: matches 200. .250 of consensus"	repeat_region	/note="Alusc repeat: matches 1. .300 of consensus"
repeat_region	11301. .11504	repeat_region	35692. .37115
repeat_region	/note="Alubd repeat: matches 85. .287 of consensus"	repeat_region	/note="L2 repeat: matches 477. .2141 of consensus"
repeat_region	11664. .11708	repeat_region	37415. .37828
repeat_region	/note="MIR repeat: matches 72. .116 of consensus"	repeat_region	/note="WSIB repeat: matches 6. .426 of consensus"
repeat_region	13586. .13665	repeat_region	38428. .38572
repeat_region	/note="40 copies 2 mer aa 64% conserved"	repeat_region	/note="MIR repeat: matches 111. .262 of consensus"
repeat_region	13784. .14056	repeat_region	38626. .39246
repeat_region	/note="L2 repeat: matches 1484. .1801 of consensus"	repeat_region	/note="LIM4 repeat: matches 4297. .4933 of consensus"
repeat_region	15463. .15706	repeat_region	39346. .39651
repeat_region	/note="LIMEC repeat: matches 2371. .2296 of consensus"	repeat_region	/note="Aluco repeat: matches 1. .301 of consensus"
repeat_region	15739. .16233	repeat_region	39699. .40087
repeat_region	/note="MIR1D repeat: matches 1. .500 of consensus"	repeat_region	/note="L2 repeat: matches 1879. .2298 of consensus"
repeat_region	16772. .17099	repeat_region	40158. .40665
repeat_region	/note="LIM1 repeat: matches 3067. .3032 of consensus"	repeat_region	/note="LIM4 repeat: matches 4796. .5332 of consensus"
repeat_region	17100. .17412	repeat_region	40666. .40992
repeat_region	/note="Alusc repeat: matches 1. .312 of consensus"	repeat_region	/note="TIGER1 repeat: matches 2078. .2418 of consensus"
repeat_region	17413. .17859	repeat_region	41012. .42354
repeat_region	/note="LIM1 repeat: matches 2633. .3068 of consensus"	repeat_region	/note="LIP12 repeat: matches 4784. .6144 of consensus"
repeat_region	17897. .18009	repeat_region	43109. .43333
repeat_region	/note="LIP18 repeat: matches 6050. .6162 of consensus"	repeat_region	/note="MIR repeat: matches 20. .148 of consensus"
repeat_region	18077. .18220	repeat_region	43333. .43694
repeat_region	/note="LIM4 repeat: matches 2352. .2163 of consensus"	repeat_region	/note="L2 repeat: matches 1188. .1584 of consensus"
repeat_region	18345. .18494	repeat_region	44121. .44349
repeat_region	/note="75 copies 2 mer tt 60% conserved"	repeat_region	/note="MER8 repeat: matches 8. .239 of consensus"
repeat_region	18506. .18646	repeat_region	44360. .44465
repeat_region	/note="LIMEC repeat: matches 1851. .1992 of consensus"	repeat_region	/note="L2 repeat: matches 2602. .2712 of consensus"
repeat_region	18654. .19000	repeat_region	44502. .44695
repeat_region	/note="LIM4 repeat: matches 1057. .1405 of consensus"	repeat_region	/note="L2 repeat: matches 2302. .2491 of consensus"
repeat_region	19414. .19509	repeat_region	45695. .45961
repeat_region	/note="48 copies 2 mer tt 63% conserved"	repeat_region	/note="Aluco repeat: matches 1. .299 of consensus"
repeat_region	20280. .20351	repeat_region	46075. .46116
repeat_region	/note="MIR1 repeat: matches 91. .141 of consensus"	repeat_region	/note="21 copies 2 mer at 79% conserved"
repeat_region	20516. .20618	repeat_region	46137. .46200
repeat_region	/note="MIR1H repeat: matches 429. .543 of consensus"	repeat_region	/note="32 copies 2 mer tt 67% conserved"
repeat_region	21226. .21528	repeat_region	46646. .46790
repeat_region	/note="Alusc repeat: matches 1. .305 of consensus"	repeat_region	/note="LIME repeat: matches 5607. .5751 of consensus"
repeat_region	21802. .21932	repeat_region	46899. .47049
repeat_region	/note="MIR repeat: matches 14. .148 of consensus"	repeat_region	47326. .47575
repeat_region	23153. .23265	repeat_region	/note="MIR repeat: matches 102. .258 of consensus"
repeat_region	/note="WER41B repeat: matches 258. .370 of consensus"	repeat_region	47845. .48314
repeat_region	24524. .24628	repeat_region	/note="LIP repeat: matches 4916. .5395 of consensus"
repeat_region	/note="MIR repeat: matches 18. .141 of consensus"	repeat_region	48303. .52827
repeat_region	24629. .24927	repeat_region	/note="LIP17 repeat: matches 1236. .5781 of consensus"
repeat_region	/note="Alusc repeat: matches 1. .299 of consensus"	repeat_region	52826. .57156
repeat_region	24928. .24986	repeat_region	/note="LIP12 repeat: matches 2. .4320 of consensus"
repeat_region	/note="MIR repeat: matches 141. .194 of consensus"	repeat_region	57157. .57376
repeat_region	25802. .25928	repeat_region	/note="Aluv repeat: matches 94. .311 of consensus"
repeat_region	/note="FLAM_C repeat: matches 2. .131 of consensus"	repeat_region	57377. .59193
repeat_region	25939. .26052	repeat_region	/note="LIP12 repeat: matches 4320. .6144 of consensus"
repeat_region	/note="L2 repeat: matches 1594. .1714 of consensus"	repeat_region	58567. .58667
repeat_region	26056. .26416	repeat_region	/note="Single clone region"
repeat_region	/note="L2 repeat: matches 1890. .2200 of consensus"	repeat_region	59220. .59596
repeat_region	26733. .27086		
repeat_region	/note="LIR40b repeat: matches 86. .446 of consensus"		

Query Match

12.0%; Score 38.4; DB 10; Length 138888;

This Page Blank (uspto)

DT 06-JAN-1999 (first entry)
DE CDNA sequence of prostate tumour clone P5105.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN M09837093-A2.
PD 27-AUG-1998.
PE 25-FEB-1998; 003492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIAX CORP.
PI Dillon DC, Xu J.
PI WPI: 98-609886/51.
PR Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3: Page 124: 130pp: English
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 320 bp: 111 A; 56 C; 64 G; 84 T;

Query Match 99.4%; Score 318; DB 1; Length 320;
Best Local Similarity 100.0%; Pred. No. 1.3e-74;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCTGAGGCTTCTGTTAGAAATAGTACAGTACAAACCAATAGAAACAAAGAA 60
DB 1 CCCCTGAGGCTTCTGTTAGAAATAGTACAGTACAAACCAATAGAAACAAAGAA 60
QY 61 AAAAGTTGTGACATGTAGTGGAGGTGTACCCCTTACTCCCATCAAAAAAAAT 120
DB 61 AAAAGTTGTGACATGTAGTGGAGGTGTACCCCTTACTCCCATCAAAAAAAAT 120
QY 121 GGATACATGTTAAAGATARAAGGCAATATTTATCATATGTTCTAAAGAGAA 180
DB 121 GGATACATGTTAAAGATARAAGGCAATATTTATCATATGTTCTAAAGAGAA 180
QY 181 GAGAAATATCTACTTCTRAAATGGAAGCCCTTAAAGTCTTTATACGAGACAC 240
DB 181 GAGAAATATCTACTTCTRAAATGGAAGCCCTTAAAGTCTTTATACGAGACAC 240
QY 241 AAATGGCCGCGCATCTCTTARAGTGCATGAGCTGGACAGGTACGTTGACG 300
DB 241 AAATGGCCGCGCATCTCTTARAGTGCATGAGCTGGACAGGTACGTTGACG 300
QY 301 TTTARACTCMGATTTGTGAC 320
DB 301 TTTARACTCMGATTTGTGAC 320

RESULT 3
V21209_01/c
Continuation (2 of 17) of V21209 from base 100001 (Methanococcus jannaschii circular chr

WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

WP	Fragment Name	Begin	End
WP	V21209_00	1	110000
WP	V21209_01	100001	210000
WP	V21209_02	200001	310000
WP	V21209_03	300001	410000
WP	V21209_04	400001	510000
WP	V21209_05	500001	610000
WP	V21209_06	600001	710000
WP	V21209_07	700001	810000
WP	V21209_08	800001	910000
WP	V21209_09	900001	1010000
WP	V21209_10	1000001	1110000
WP	V21209_11	1100001	1210000
WP	V21209_12	1200001	1310000
WP	V21209_13	1300001	1410000

WP V21209_14 1400001 1510000
WP V21209_15 1500001 1610000
WP V21209_16 1600001 1664976

Query Match 14.1%; Score 45; DB 1; Length 110000;
Best Local Similarity 48.2%; Pred. No. 0.0093;
Matches 120; Conservative 2; Mismatches 127; Indels 0; Gaps 0;

QY 5 TGAAGCTTCTGTAGAAATAGTACAGTACCAACCAATAGAAACAAAGAA 64
DB 93590 TGAAGCAACCTTAAAGATGAAGAAATATTTAAACCAATATCCAAACATATATAAAA 93591
QY 65 GTTGTGACATGTAGTGGAGGTGTACCCCTTACCCTCCATCAAAAAAAATGAT 124
DB 93530 ATGTGTTGAAGTTTAAAGACAGATGAGATATGATTCCTCAAAAAACAAAAATGGTT 93471
QY 125 ACATGTTAAAGATARAAGGCAATATTTATCATATGTTCTAAAGAGAAAGAGA 184
DB 93470 TATGTCGACGATGAGAACTATATAATTTTATATAAAATCTAAATAAAACGAAT 93411
QY 185 AATATCTACTTCTRAAATGGAAGCCCTTAAAGTCTTTGATCTGAGACCAAT 244
DB 93410 ATATGTTCCACTTATGACAGAAACCATATATAGTTTATAGGATGATGAACT 93351
QY 245 GTGCCGTC 253
DB 93350 ATTACTTC 93342

RESULT 4
V21209_15/c
Continuation (16 of 17) of V21209 from base 1500001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

WP	Fragment Name	Begin	End
WP	V21209_00	1	110000
WP	V21209_01	100001	210000
WP	V21209_02	200001	310000
WP	V21209_03	300001	410000
WP	V21209_04	400001	510000
WP	V21209_05	500001	610000
WP	V21209_06	600001	710000
WP	V21209_07	700001	810000
WP	V21209_08	800001	910000
WP	V21209_09	900001	1010000
WP	V21209_10	1000001	1110000
WP	V21209_11	1100001	1210000
WP	V21209_12	1200001	1310000
WP	V21209_13	1300001	1410000
WP	V21209_14	1400001	1510000
WP	V21209_15	1500001	1610000
WP	V21209_16	1600001	1664976

Query Match 11.6%; Score 37; DB 1; Length 110000;
Best Local Similarity 49.2%; Pred. No. 1.1;
Matches 94; Conservative 1; Mismatches 96; Indels 0; Gaps 0;

QY 56 AAGAAGAAAGTTGTGACATTTAGTAGGAGGTGTACCCCTTACTCCCAATCAAAAA 115
DB 61937 AAGAAGAAAGTTTAAATCTTAAAGGTTTATATATAGTAGTACTTACCTTGA 61878
QY 116 AATAGATACATGTTAAAGATARAAGGCAATATTTATCATATGTTCTAAAGAA 175
DB 61877 ATATGTTATGAAAGCTTAAATATTTGCGGAGTTGAAGTATATATGTTGAAATG 61818
QY 176 AGAAGAGAAATATCTTCTCRAAATGGAAGCCCTTAAAGTCTTTGATGATGAG 235
DB 61817 GCCCATTTAAATTAAGACCGTTGCGAATGGAATCGTAAATGTTGCTTATGACT 61758
QY 236 GACACCAATGT 246
DB 61757 AAAACACAGT 61747


```

RESULT 5
ID T72882
AC T72882 standard: cDNA: 19124 BP.
DT 12-SEP-1997 (first entry)
DE Plasmodium var-7 gene.
KM DBL gene family: SABP: stailic acid binding protein; vaccine; therapy;
KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
KM DABP; mezozyote; malaria; var-1; var-2; var-3; var-7; immune response;
OS Plasmodium vivax.
OS Plasmodium falciparum.
FH Key location/Qualifiers
FT exon 7317..15139
FT FT /*tag= a
FT FT /number= 1
FT FT 15140..16205
FT FT /*tag= b
FT FT /number= 1
FT FT 16206..17552
FT FT /*tag= c
FT FT /number= 2
FT FT /note= "no stop codon given"
PN MO9640766-A2.
PD 19-DEC-1996.
PF 07-JUN-1996: U09508
PR 07-JUN-1995: US-487826.
PI (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Chitlins C, Miller LH, Peterson DS, Sim KL, Su X;
PI Wellens TE;
PI WPI: 97-052231/05.
DR P-PSDB: W22475.
DE New malaria vaccines - contains cysteine-rich DBL family protein
PT binding domains homologous domains of the Duffy and stailic acid
PT binding proteins
PS Claim 4: Page 56-61: 96pp: English.
CC This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to
CC the Duffy binding like (DBL) family of genes which have homology to the
CC Duffy antigen binding protein (DABP) and stailic acid binding protein
CC (SABP) conserved regions (see T72889 and T72888 respectively). The var
CC family of genes modulate cytoadherence and antigenic variation of
CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC protein (DABP) are soluble proteins that appear in the culture
CC supernatant after infected erythrocytes release merozoites. DABP and SABP
CC mediate the binding of merozoites and schizonts to the erythrocyte
CC surface. These proteins are necessary for erythrocyte invasion by the
CC parasite. This sequence can be used in the compositions of malaria, and
CC The compositions are for the treatment and prevention of malaria, and
CC comprise either a nucleotide sequence or encoded polypeptide of the
CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
CC genes having homology with conserved regions of DABP and SABP. The
CC compositions are used for the treatment and prevention of malaria. They
CC are also used in the preparation of vaccines for inducing a protective
CC immune response in a mammal to Plasmodium merozoites (especially
CC Plasmodium falciparum or Plasmodium vivax).
SQ Sequence 19124 BP: 7824 A; 2190 C; 2790 G; 6320 T;

```

```

Query Match 11.4%; Score 36.6; DB 1; Length 19124;
Best Local Similarity 51.5%; Pred. No. 0.96;
Matches 84; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

```

```

OY 36 ACAACCAATAGCAACACAAAGAAAGTTGTGACATTGTAGTACGAGGTGTGATACC 95
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 402 ACAACCAATAGCAACACAAAGAAAGTTGTGACATTGTAGTACGAGGTGTGATACC 461
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 96 CCTTACTGCCATGCAAAATGATGATGTTAAAGATTAAGGATTAAGGATTAAGT 155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 462 CATACGTCATTTTATTAATTAATGATTTATTCGATTAAGCAATTAATGATTAAT 521
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 156 ATCATATGTTCTAAAGAGAAAGAAAGAAATACTACTTCT 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 522 AATTAATTTACATAGAAAGATTAAAGATTAACATTTT 564
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 6
ID T68041
AC T68041 standard: DNA: 1215 BP.
DT 21-JUL-1997 (first entry)
DE H. pylori cytoplasmic protein ORF 07gp1807orf42.
KM Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KM duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
OS Helicobacter pylori.
FH Key location/Qualifiers
FT cds 1..1215
FT FT /*tag= a
FT FT /note= "no stop codon given"
PN MO9640893-A1.
PD 19-DEC-1996.
PF 06-JUN-1996: U09122.
PR 07-JUN-1995: US-487032.
PR 01-APR-1996: US-630405.
PI (ASTR ) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaerd BL;
PI WPI: 97-052306/05.
DR P-PSDB: W20788.
DE Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
PS Claim 9: Page 862: 1481pp: English.
CC This protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The genomic sequence of H. pylori (ATCC 55679) was determined from
CC overlapping contigs generated by mechanically shearing the bacterial
CC DNA. The sequences were analysed for ORF of at least 180 nucleotides.
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
SQ Sequence 1215 BP: 379 A; 254 C; 214 G; 368 T;

```

```

Query Match 11.1%; Score 35.6; DB 1; Length 1215;
Best Local Similarity 56.0%; Pred. No. 0.93;
Matches 65; Conservative 1; Mismatches 50; Indels 0; Gaps 0;

```

```

OY 128 TGGTAAAGATTAAGGCAATTTTATCATATGTTCTAAAGAGAAAGAGAGAAA 187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 857 TGGTAAAGGCAACGCTGATTTATTTTAAAGAAATGAAGAAAT 916
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 188 TACTACTTTCTCRAAATGAAGCCCTTAAGGTCCTTGATATCAGAGACACAAA 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 917 TCATCATTTTTCATGCGAAGCCCTTAAGCCTTAAGAGCGATGCGTTAAA 972
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 7
ID T67711
AC T67711 standard: DNA: 717 BP.
DT 17-JUL-1997 (first entry)
DE H. pylori cytoplasmic protein ORF 719606_aa.
KM Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KM duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
OS Helicobacter pylori.
FH Key location/Qualifiers
FT cds 1..717

```

```
FT      /tag= a
FT      /note= "no stop codon given"
PN      WO9640893-A1.
PD      19-DEC-1996.
PF      06-JUN-1996; U09122.
PR      07-JUN-1996; US-487032.
PR      01-APR-1996; US-630405.
PA      (ASTR) ASTRA AB.
PI      Berglinch OT, Smith D, Mellgaard BL;
DR      WPI: 97-052306/05.
DR      P-PSDB: W20567.
PT      Helicobacter pylori nucleic acid sequences and related
PT      polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT      infection, and to detect Helicobacter
PS      Claim 9; Page 315-318; 1481pp; English.
CC      This sequence encodes a H. pylori cytoplasmic protein.
CC      The protein may be used in a vaccine to prevent or treat H. pylori
CC      infection or to identify H. pylori polypeptide binding compounds,
CC      useful as potential H. pylori life cycle activators or inhibitors.
CC      The genomic sequence of H. pylori (ATCC 55679) was determined from
CC      overlapping contigs generated by mechanically shearing the bacterial
CC      DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC      and the predicted coding regions defined by computer evaluation. To
CC      identify likely H. pylori antigens for vaccine development, the amino
CC      acid sequences predicted from various ORF were analysed for significant
CC      homology to other known or exported membrane proteins. Having identified
CC      and determined the sequences of interest, particular regions can be
CC      isolated from H. pylori by PCR amplification for recombinant polypeptide
CC      production, e.g. in E. coli hosts.
SQ      Sequence 717 BP; 224 A; 127 C; 146 G; 220 T;

Query Match
Best Local Similarity 11.1%; Score 35.6; DB 1; Length 717;
Matches 65; Conservative 1; Mismatches 50; Indels 0; Gaps 0;

QY      128 TGGTAAAGATARAAGGCAATATTTATCATATGTTCTAAAGAGAGAGAGAAA 187
DB      434 TGGGTAATGGGCAACCGATTTATTTAGATTTTAAAGAAATGAAGAAAAT 493
QY      188 TACTATTTCTCAAAATGGAAGCCCTTAAGAGTCCTTGATCTGAAGACACAAA 243
DB      494 TCATCATTTTTCATGCGGAACGCTTTAAAGCCTTTAAAGCGCATGGCGTTAA 549

RESULT 8
V7949
ID      V7949 standard; DNA: 1107 BP.
AC      V7949;
DE      16-MAR-1999 (first entry)
KW      Staphylococcus aureus contig SEQ ID #638.
KW      Computer readable medium; vaccine; S. aureus infection; immunodetection;
KW      cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW      skin infection; surgical wound infection; scalded skin syndrome;
KW      toxic shock syndrome; ds.
OS      Staphylococcus aureus.
FH      Key
FT      Location/Qualifiers
FT      misc_feature
FT      541..600
FT      /tag= a
FT      /note= "these bases represent a line of missing text in
FT      the sequence listing in the specification. They
FT      are included to maintain the nucleotide numbering
FT      given in the specification for this DNA sequence"
EP      786519-A2.
PD      30-JUL-1997.
PF      07-JAN-1997; 100117.
PR      05-JAN-1996; US-009861.
PA      (HUMA-) HUMAN GENOME SCI INC.
PI      Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI      Rosen CA;
DR      WPI: 97-374922/35.
DR      Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT      stored on computer readable medium and used in the production of
```

```
PT      anti-S. aureus vaccines
PS      Claim 1; Page 1565-1566; 3271pp; English.
CC      This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC      of the invention. The DNA sequences are recorded on a computer readable
CC      medium, preferably selected from a floppy or hard disk, random access
CC      memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC      the S. aureus DNA sequences allows putative functions to be assigned so
CC      that protein-encoding or regulatory regions of commercial, therapeutic or
CC      industrial importance can be obtained. Specifically, sequences which are
CC      likely to encode antigens have been identified and these polypeptides can
CC      be used in a vaccine composition against S. aureus infection. The
CC      polypeptides can also be used in a kit for the immunodetection of
CC      S. aureus in a sample. S. aureus is implicated in numerous human diseases,
CC      including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC      skin and surgical wound infections, scalded skin syndrome, toxic shock
CC      syndrome, etc. Organisms transformed with the DNA sequences can be used
CC      for recombinant production of the polypeptides. The new DNA sequences
CC      (and their fragments) are useful as primers or probes for isolating
CC      homologues of any of the S. aureus DNA sequences contained on the
SQ      computer readable medium.
SQ      Sequence 1107 BP; 384 A; 134 C; 194 G; 329 T;

Query Match
Best Local Similarity 10.9%; Score 35; DB 1; Length 1107;
Matches 56; Conservative 2; Mismatches 39; Indels 0; Gaps 0;

QY      108 TCAAAAAAATGATACATGCTTAAGATARAAGGCAATATTTATCATGTTCT 167
DB      221 TAAAAATAAAAATGGGGAGATTATATGCTTAACATTCAGAGCTTANCTACTTCT 280
QY      168 AAAAGAGAGAGAGAGAAATACTACTCTTCRAAT 204
DB      281 AAAAGAGAGAGAGAGAAATCAATATACCGTTTCGTAGAT 317

RESULT 9
V17183
ID      V17183 standard; DNA: 1811 BP.
AC      V17183;
DE      18-JUN-1998 (first entry)
KW      Insecticidal toxin subunit BrhTX-1(a)1.1 genomic sequence.
KW      Insecticidal toxin; Bracon hebetor; insect control; pathogen;
KW      recombinant baculovirus; ds.
OS      Bracon hebetor.
FH      WO9744355-A1.
PD      27-NOV-1997.
PF      01-MAY-1997; G01205.
PR      22-MAY-1996; GB-010748.
PR      22-MAY-1996; GB-010687.
PR      22-MAY-1996; GB-010695.
PR      22-MAY-1996; GB-010697.
PR      22-MAY-1996; GB-010738.
PR      22-MAY-1996; GB-010739.
PA      (CSIR) COMMONWEALTH SCI & IND RES ORG.
PA      (ZENE) ZENECA LTD.
PI      Christian PD, Claudianos A, Daly A, Duncan RE, Suner M,
PI      Windass JD;
DR      WPI: 98-018430/02.
PT      New nucleic acid encoding a combination of insecticidal sub-unit(s)
PT      of wasp toxin - and related transformed cells, insect pathogens and
PT      combinations of proteins, useful as insecticides
PS      Claim 1; Pages 57-58; 84pp; English.
CC      This is a contiguous genomic sequence which encodes part of a spliced RNA
CC      that is related to the cDNA comprised by pBrhTX-1(a)1.1. This 1811 base
CC      pair spliced RNA derived from the Bracon hebetor genomic clone encodes
CC      at least two of the insecticidal toxin subunits shown in sequences
CC      W52124-W52128. The spliced RNA can hybridise with extension products
CC      prepared from a 564 base pair (V17145) template with 6 specified primers.
CC      A nucleic acid encoding at least one of the specified subunits can be
CC      modified so that mRNA instability motifs and/or fortuitous splice sites
CC      are removed, or insect-pest preference codons are used, so that
CC      expression of this nucleic acid in insect cells yields practically the
```


Query Match 10.7%; Score 34.2; DB 1; Length 110000;
Best Local Similarity 47.8%; Pred. No. 6.1;
Matches 96; Conservative 1; Mismatches 104; Indels 0; Gaps 0;

QY 12 TTCTGTGTAAGAAATAGTACAGTTACACCAATAGACACAAAGAAAAGTTGTG 71
DB 39340 TCGTTGAAATACATATTAAATAGAAACACAAAGACTTTGAAACCTCG 39399
QY 72 ACATTGTAGGAGGAGTGTACCCCTTACCCATCAAAAAAATGATCATGTGT 131
DB 39400 CATTGTAGCAATTAATTTTATTTTATTTTATTAATAAAGTCCAGCATAG 39459
QY 132 TAAAGATARAAGGCAATTTTATCATATGTTCTAAAGAGAGAGAAATACT 191
DB 39460 TAATTAATTAATTAATTTTATTAATTAATTAATTAATTAATTAATAA 39519
QY 192 ACTTCTCRAAATGAGAGCC 212
DB 39520 AATATCTCAAAATTTAACTC 39540

RESULT 12
X20249/c
ID X20249 standard; DNA; 116277 BP.
AC X20249;
DE 04-MAY-1999 (first entry)
DE Borrelia burgdorferi polynucleotide sequence #2.
KW Borrelia burgdorferi; Spirochete; bacterium; pathogen; Lyme disease;
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW infection; diagnosis; characterisation; detection; ds.
OS Borrelia burgdorferi.
PN WO9858943-A1.
PD 30-DEC-1998.
PF 18-JUN-1998; U12764.
PR 03-SEP-1997; US-057483.
PR 20-JUN-1997; US-050359.
PR 22-JUL-1997; US-053344.
PR 22-JUL-1997; US-053377.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI White OR;
PI WPI: 99-081217/07.
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
PS Claim 1; Page 672-737; 1128pp; English.
CC X20248 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
SQ Sequence 116277 BP; 42656 A; 19868 C; 14490 G; 39250 T;

Query Match 10.7%; Score 34.2; DB 1; Length 116277;
Best Local Similarity 47.8%; Pred. No. 6.2;
Matches 96; Conservative 1; Mismatches 104; Indels 0; Gaps 0;

QY 12 TTCTGTGTAAGAAATAGTACAGTTACACCAATAGACACAAAGAAAAGTTGTG 71
DB 115794 TCGTTGAAATACATATTAAATAGAAACAAACAGAACTTTGAAACCTCG 115735
QY 72 ACATTGTAGGAGGAGTGTACCCCTTACCCATCAAAAAAATGATCATGTGT 131
DB 115734 CATTGTAGCAATTAATTTTATTTTATTTTATTAATAAAGTCCAGCATAG 115675

QY 132 TAAAGATARAAGGCAATTTTATCATATGTTCTAAAGAGAGAAATACT 191
DB 115674 TAATTAATTAATTAATTTTATTAATTAATTAATTAATTAATAA 115615
QY 192 ACTTCTCRAAATGAGAGCC 212
DB 115614 AATATCTCAAAATTTAACTC 115594

RESULT 13
X20259/c
ID X20259 standard; DNA; 9842 BP.
AC X20259;
DE 04-MAY-1999 (first entry)
DE Borrelia burgdorferi polynucleotide sequence #12.
KW Borrelia burgdorferi; Spirochete; bacterium; pathogen; Lyme disease;
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW infection; diagnosis; characterisation; detection; ds.
OS Borrelia burgdorferi.
PN WO9858943-A1.
PD 30-DEC-1998.
PF 18-JUN-1998; U12764.
PR 03-SEP-1997; US-057483.
PR 20-JUN-1997; US-050359.
PR 22-JUL-1997; US-053344.
PR 22-JUL-1997; US-053377.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI White OR;
PI WPI: 99-081217/07.
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
PS Claim 1; Page 914-919; 1128pp; English.
CC X20248 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
SQ Sequence 9842 BP; 3671 A; 1321 C; 1151 G; 3699 T;

Query Match 10.6%; Score 34; DB 1; Length 9842;
Best Local Similarity 46.6%; Pred. No. 3.9;
Matches 97; Conservative 3; Mismatches 108; Indels 0; Gaps 0;

QY 99 TACTCCCATCAAAAAAATGATACATGTTAAAGATARAAGGCAATTTTATC 158
DB 6937 TATGTGCAATTAATTAATAAATTAATTAATTAATTAATTAATTAATTAAT 6878
QY 159 ATATGTTTAAAGAGAGAGAGAAATACTACTTCTCAAAATGAGAGCCCTTAAG 218
DB 6877 ACAATTTAAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6818
QY 219 GTGCTTGAATGAGAGAGACAAATGTGGCGGTCATCTCTTARAGTTGATGACT 278
DB 6817 AAAAGACAGAAAAAACCCACTATTAGAGAGTCGATACGGCGGAGATGCAAAAT 6758
QY 279 TGGACAGGTAATGATGATTAATTAATTAATTAATTAATTAATTAATTAAT 306
DB 6757 TGTAGATGATCCTGATGAGCAATAGAA 6730

RESULT 14
X20248_03
Continuation (4 of 10) of X20248 from base 300001 (Borrelia burgdorferi polynucleotide
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248
Fragment Name Begin End

WP X20248_00 1 110000
 WP X20248_01 100001 210000
 WP X20248_02 200001 310000
 WP X20248_03 300001 410000
 WP X20248_04 400001 510000
 WP X20248_05 500001 610000
 WP X20248_06 600001 710000
 WP X20248_07 700001 810000
 WP X20248_08 800001 910000
 WP X20248_09 900001 910715

Query Match 10.4%; Score 33.8; DB 1; Length 110000;
 Best Local Similarity 53.4%; Pred. No. 7.8;
 Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 42 AATGAGAACACAAAGAAAGATTGTGACATTGTGTGAGGAGTGTGACCCCTTAC 101
 DB 14419 AATATCTCAACAAATMAAAGAAACCTTACCAATMAAAGCTTCTAAATCTCCCA 14478
 QY 102 TCCCATCAAAAAAATGATGATCATGTTAAGGATARAAGGCAATTTTATCAT 161
 DB 14479 ACTTCACACAAATGAAATTAATATGATAGAGCTTATATTCCTCAATAGCCCAACTT 14538
 QY 162 TGTCTTAAAGAG 174
 DB 14539 CTTTCAAAAAATAG 14551

RESULT 15
 ID T72883
 DT T72883 standard; cDNA; 7295 BP.
 DE 12-SEP-1997 (first entry)
 KW DBL gene family; SABP; stialic acid binding protein; vaccine; therapy;
 KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 OS Plasmodium vivax.
 OS Plasmodium falciparum.
 FH Key location/Qualifiers
 FT cds 517..7065
 FT /*tag- a
 FT /product- var-1
 PN W09640766-A2.
 PD 19-DEC-1996.
 PR 07-JUN-1996; 009508.
 PR 07-JUN-1995; US-487826.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Chittins C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 DR WPI: 97-052231/05.
 DR P-PSDB: W22476.
 PT New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and stialic acid
 PT binding proteins
 PS Claim 4; Page 67-69; 96pp; English.
 CC This sequence represents the var-1 gene of Plasmodium. Var-1 belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and stialic acid binding protein
 CC (SABP) conserved regions (see T72889 and T72888 respectively). The var
 CC family of genes modulate cytoadherence and antigenic variation of
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 CC protein (DABP) are soluble proteins that appear in the culture
 CC supernatant after infected erythrocytes release merozoites. DABP and SABP
 CC mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the invention.
 CC The compositions are for the treatment and prevention of malaria, and
 CC comprise either a nucleotide sequence or encoded polypeptide of the
 CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
 CC genes having homology with conserved regions of DABP and SABP. The

CC compositions are used for the treatment and prevention of malaria. They
 CC are also used in the preparation of vaccines for inducing a protective
 CC immune response in a mammal to Plasmodium merozoites (especially
 CC Plasmodium falciparum or Plasmodium vivax).
 SQ Sequence 7295 BP; 2866 A; 1067 C; 1455 G; 1907 T;

Query Match 10.4%; Score 33.2; DB 1; Length 7295;
 Best Local Similarity 51.4%; Pred. No. 5.9;
 Matches 74; Conservative 1; Mismatches 69; Indels 0; Gaps 0;

QY 15 TTGTAGAAATAGTACAGTTACCAACCAATAGCAACAAAGAAAGTTGTGACA 74
 DB 4885 TGCTAGAAATATTTTATAGAGATTATATGAAATTAAGAAATTAAGCTATGACG 4944
 QY 75 TTGTAGTAGGAGGTGTACCCCTTACTCCCATCAAAAAAATGATCATGTTAA 134
 DB 4945 AAAAAGGAAGATGATGCAATGATTAAGAGTTGTATAGAAAAATGGTACAGAAAA 5004
 QY 135 AGATARAAGGCAATATTTATC 158
 DB 5005 ACGAAGATGCAAAAAATTAAC 5028

Search completed: September 25, 1999, 09:58:52
 Job time: 8183 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 07:30:14 ; Search time 165.01 Seconds
(without alignments)
177.786 Million cell updates/sec

Title: US-09-030-606-224

Perfect score: 320
Sequence: 1 CCCCTGAGGCTCTTGTGTA.....TTTAACTGCGATTGTGAC 320

Scoring table: IDENTITY_NUC

Searched: 176461 seqs, 45838279 residues

Database:

Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/1na/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/1na/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/1na/5C_COMB.seq: *
4: /cgn2_6/ptodata/2/1na/5D_COMB.seq: *
5: /cgn2_6/ptodata/2/1na/PTUS9_COMB.seq: *
6: /cgn2_6/ptodata/2/1na/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38.2	11.9	7218	2	US-08-232-463-14
2	34.4	10.8	2367	3	US-08-441-139-3
3	31.8	9.9	1920	3	US-08-720-899-1
4	31.8	9.9	1920	3	US-08-459-610-1
5	31.8	9.9	1920	4	US-08-343-804-1
6	31.6	9.9	5852	1	US-07-867-106-2
7	31	9.7	4059	4	US-08-485-139-1
8	30.6	9.6	4435	2	US-08-308-872B-5
9	30.6	9.6	321	4	US-08-520-678A-23
10	30.4	9.5	9636	3	US-08-323-170B-1
11	30.2	9.4	19557	5	PCT-US92-06300-1
12	29.8	9.3	12124	1	US-08-181-271A-36
13	29.8	9.3	12124	2	US-08-449-315-36
14	29.8	9.3	12124	2	US-08-444-803-36
15	29.8	9.3	12124	2	US-08-449-043-36
16	29.8	9.3	12124	3	US-08-456-265A-36
17	29.8	9.3	12124	3	US-08-455-416-36
18	29.8	9.3	12124	3	US-08-455-244-36
19	29.8	9.3	12124	3	US-08-454-876-36
20	29.8	9.3	12124	4	US-08-457-364-36
21	29.8	9.3	12124	4	US-08-456-262-36
22	29.8	9.3	12124	4	US-08-456-240-36
23	29.8	9.3	12124	4	US-08-455-736-36
24	29.6	9.2	15108	1	US-08-157-005-1
25	29.6	9.2	15101	5	PCT-US95-09927-14
26	29.4	9.2	1824	1	US-08-347-792-5
27	29.4	9.2	1824	2	US-08-431-357-5
28	29.4	9.2	2277	3	US-08-676-967-2
29	29.4	9.2	2291	3	US-08-355-888A-6
30	29.4	9.2	2277	3	US-08-676-974-2
31	29.4	9.2	2606	4	US-08-568-459A-7
32	29.4	9.2	2991	4	US-08-588-190-1
33	29.4	9.2	2991	4	US-08-693-697-6
34	29.4	9.2	2880	4	US-08-693-697-32
35	29.4	9.2	2877	4	US-08-693-697-35
36	29.4	9.2	246240	4	US-08-724-394A-20
37	29.4	9.2	246240	4	US-08-724-394A-21

38	29.4	9.2	246240	4	US-08-724-394A-22	Sequence 22, Appl
39	29.4	9.2	2880	4	US-08-588-526-1	Sequence 1, Appl
40	29.4	9.2	1824	5	PCT-US95-15353-5	Sequence 5, Appl
41	29	9.1	5181	3	US-08-257-073-10	Sequence 10, Appl
42	29	9.1	11832	4	US-08-416-603-1	Sequence 1, Appl
43	29	9.1	356	4	US-08-520-678A-22	Sequence 22, Appl
44	29	9.1	257	4	US-08-520-678A-24	Sequence 24, Appl
45	29	9.1	253	4	US-08-520-678A-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1

US-08-232-463-14/C

Sequence 14, Application US/08232463

Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: DORNER, F.

APPLICANT: SCHEIFLINGER, F.

APPLICANT: FALKNER, F. G.

TITLE OF INVENTION: RECOMBINANT FOXLPOX VIRUS

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 1800 Diagonal Road, Suite 500

CITY: Alexandria

STATE: VA

COUNTRY: USA

ZIP: 22313-0299

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/232,463

FILING DATE:

CLASSIFICATION: 435

Prior Application DATA:

APPLICATION NUMBER: US/07/935,313

FILING DATE:

APPLICATION NUMBER: EP 91 114 300.6

FILING DATE: 26-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/114 IMMU

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)836-9300

TELEFAX: (703)683-4109

TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 7218 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: pTZgt-Fls

US-08-232-463-14

Query Match 11.9%; Score 38.2; DB 2; Length 7218;

Best Local Similarity 8.7%; Pred. No. 0.033;

Matches 24; Conservative 138; Mismatches 113; Indels 0; Gaps 0;

20 AGAAATAGTACAGTACCAACCAATAGGACACAAAGAAAGTGTGACATGTA 79

1289 RRR 1230

80 GTAGGAGTGTGTACCCCTTACTCCCATCAAAAAAATGATACATGCTTAAGCAT 139


```

: Patent No.5830837
:
: GENERAL INFORMATION:
: APPLICANT: Bisgaard-Frantzen, Henrik
: APPLICANT: Borchert, Torben Vedel
: APPLICANT: Svendsen, Allan
: APPLICANT: Thellersen, Marianne
: APPLICANT: Van der Zee, Pia
: TITLE OF INVENTION: AMYLASE VARIANTS
: NUMBER OF SEQUENCES: 38
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 5830837c No. 5830837disk of No. 5830837th America, Inc.
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10174-6401
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/343,804
: FILING DATE: 22-NOV-1994
:
: CLASSIFICATION: 435
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Lowmey Dr., Karen A.
: REGISTRATION NUMBER: 31,274
: REFERENCE/DOCKET NUMBER: 4054, 214-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
:
: INFORMATION FOR SEQ ID NO: 1:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1920 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 334..1872
:
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 334..420
:
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: 421..1869
:
: US-08-343-804-1
:
: Query Match 9.9%; Score 31.8; DB 4; Length 1920;
: Best Local Similarity 56.8%; Pred. No.1.4; Mismatches 39; Indels 0; Gaps 0;
: Matches, 54; Conservative 2;
:
: Oy 110 AAAAAAAAAATGATACATGTTAAAGATAPAAAGCGCAATTTTATCATATGTTCTAA 169
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: DB 187 AGAAGCAGAGAGGATATGAATTAATGATAGTAAGACGCGCATATGCGGCTTTCTTTGG 246
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: Oy 170 AAGAGAGAGAGAGAAAATACTACTTCTTCRAAT 204
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
: DB 247 AAGAAAAATATAGGAAAATGACTTGTAAAAAT 281
: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
:
: RESULT 5
: Sequence 2, Application US/07867106
: Patent No. 5389526
:
: GENERAL INFORMATION:
: APPLICANT: Ståde, Martin B
: APPLICANT: Chang, Andy C M
: APPLICANT: Williams, Keith L
: TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
: Slime Moulds of the Genus Dictyostelium

```

NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: AU PJ 7187
FILING DATE: 02-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Feeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
FEATURE:
NAME/KEY: CDS
LOCATION: 2378..5038
US-07-867-106-2

Query Match 9.9%; Score 31.6; DB 1; Length 5852;
Best Local Similarity 51.5%; Pred. No. 2.4;
Matches 70; Conservative 1; Mismatches 65; Indels 0; Gaps 0;

QY 36 ACACCAATGAGCAACAAAAAGTTGTGACATGTGATGAGGAGTGTACC 95
DB 5449 AAAAAAAAAAAAAAAAAAATCAATATGTTATGTTTAAAGATTTTAAATC 5508
QY 96 CCTTACCCCATCAAAAAAATGATGATCATGTTAAAGATARAAGGCAATATTT 155
DB 5509 TCGTCATGATGTTTAAATATGATCATATTTTAAAAAACCCTTACATTTT 5568
QY 156 ATCATATGTTCTAAA 171
DB 5569 TTATTTAATTCAAA 5584

RESULT 7
US-08-485-139-1/c
Sequence 1, Application US/08485139
Patent No. 5880331
GENERAL INFORMATION:
APPLICANT: KREBBERS, Enno
APPLICANT: WILLIAMS, Mark
APPLICANT: LEMANS, Jan
TITLE OF INVENTION: USE OF ANTHOCYANIN GENES TO MAINTAIN
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:

ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,139
FILING DATE: 07-JUN-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm R.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-096
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4059 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: C1 gene of Zea mays
FEATURE:
NAME/KEY: -
LOCATION: 279..284
OTHER INFORMATION: /label= HpaI
FEATURE:
NAME/KEY: -
LOCATION: 447..452
OTHER INFORMATION: /label= EcoRI
FEATURE:
NAME/KEY: -
LOCATION: 1735..1740
OTHER INFORMATION: /label= AatII
FEATURE:
NAME/KEY: -
LOCATION: 1505..1510
OTHER INFORMATION: /label= EcoRI
FEATURE:
NAME/KEY: -
LOCATION: 2081..2086
OTHER INFORMATION: /label= XhoI
FEATURE:
NAME/KEY: -
LOCATION: 2418..2430
OTHER INFORMATION: /label= SfiI
FEATURE:
NAME/KEY: -
LOCATION: 2659..2674
OTHER INFORMATION: /label= SnaBI
FEATURE:
NAME/KEY: -
LOCATION: 3008..3013
OTHER INFORMATION: /label= HpaI
FEATURE:
NAME/KEY: -
LOCATION: 1..1077
OTHER INFORMATION: /label= Pci
OTHER INFORMATION: /note= "region containing promoter of C1 gene"
FEATURE:

NAME/KEY: -
LOCATION: 1078..2134
OTHER INFORMATION: /label= C1
OTHER INFORMATION: /note= "coding region of C1 gene"
FEATURE:
NAME/KEY: -
LOCATION: 2135..2430
OTHER INFORMATION: /label= 3'C1
OTHER INFORMATION: /note= "region containing polyadenylation signal"
OTHER INFORMATION: of C1 gene"
FEATURE:
NAME/KEY: -
LOCATION: 1033..1038
OTHER INFORMATION: /label= TATA-Box
FEATURE:
NAME/KEY: -
LOCATION: 1061..1062
OTHER INFORMATION: /label= transcript-init
OTHER INFORMATION: /note= "transcription initiation site"
FEATURE:
NAME/KEY: Intron
LOCATION: 1211..1299
FEATURE:
NAME/KEY: Intron
LOCATION: 1430..1575
FEATURE:
NAME/KEY: -
LOCATION: 935..939
OTHER INFORMATION: /label= C1-S
OTHER INFORMATION: /note= "TGCAG sequence (in C1 gene) which in the
OTHER INFORMATION: C1-S sequence is changed to TTAGC"
US-08-485-139-1

Query Match 9.7%; Score 31; DB 4; Length 4059;
Best Local Similarity 61.7%; Pred. No. 3.1;
Matches 66; Conservative 0; Mismatches 40; Indels 1; Gaps 1;

QY 18 TTGAAATATGACAGTACCAACCAATGAGAACCAAAAAAGTTGTGACATG 77
DB 3636 TTATAAATGACACATTTTCATCATATAGAGACAA-AATAGAGAAAGTTGTGAGATCA 3578
QY 78 TAGTAGGAGGTGTACCCCTTACTCCCATCAAAAAAATGGAT 124
DB 3577 AGTTTAAATAGTCCCATGAGATCAAACTATATAAAAAAGATAGAT 3531

RESULT 8
US-08-308-872B-5/c
Sequence 5, Application US/08308872B
Patent No. 5661006
GENERAL INFORMATION:
APPLICANT: BROWN, Thomas David Kay
TITLE OF INVENTION: HORSBURGH, Brian Colin
TITLE OF INVENTION: CANINE CORONAVIRUS SUBUNIT VACCINE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Akzo No. 5661006el Patent Department
STREET: 1300 Piccard Drive, Suite 206
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/308,872B
FILING DATE: 19-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,846

FILING DATE: 05-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,641
FILING DATE: 24-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 91.303.737.0
FILING DATE: 25-APR-1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gormley, Mary E.
REGISTRATION NUMBER: 34,409
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4435 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Canine corona virus
STRAIN: CCV-V54
FEATURE:
NAME/KEY: CDS
LOCATION: 60..4418
OTHER INFORMATION: /label= CCV-C54_Spikegene
US-08-308-872B-5

Query Match 9.6%; Score 30.6; DB 2; Length 4435;
Best Local Similarity 47.5%; Pred. No. 4.1;
Matches 87; Conservative 1; Mismatches 95; Indels 0; Gaps 0;

QY 2 CCCCTAGAGCTCTGTGAGAAATGATACAGTTACACCAATGAGAACCAAAAAAGAA 61
DB 429 CTTATAGAGCGATGTGTAACATGATACAGTACGAGACTACCAAGAACATGACTAGTA 370
QY 62 AAAGTTGTGACATGTAGTAGGAGGTGTACCCCTTACTCCCATCAAAAAAATG 121
DB 369 AAGTTTACCTCTTCATTCGACAGTACATTCGCCATAGCTTCATATCAAAATTAATG 310
QY 122 GATACATGCTTAAGGATARAAGGCAATATTATCTATAGTTCTAAAGAGAAAGAG 181
DB 309 CATGATGTACTAAATATGATAGCTGATGTTGTCTGTCTGGAAGAGTTGTACC 250
QY 182 AGA 184
DB 249 ACA 247

RESULT 9
US-08-520-678A-23/c
Sequence 23, Application US/08520678A
Patent No. 5874565
GENERAL INFORMATION:
APPLICANT: Rice, Charles M.
TITLE OF INVENTION: Kolyhalov, Alexander A.
TITLE OF INVENTION: NOVEL 3' TERMINAL SEQUENCE OF HEPATITIS
C VIRUS GENOME AND DIAGNOSTIC AND THERAPEUTIC USES THEREOF
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howell & Haferkamp, L.C.
STREET: 7733 Forsyth Blvd., Suite 1400
CITY: St. Louis
STATE: MO
COUNTRY: USA
ZIP: 63105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

RESULT 10
US-08-323-1708-1
Sequence 1, Application US/083231708
Patent No. 5733772
GENERAL INFORMATION:
APPLICANT: Williamson, Kim C.
APPLICANT: Kaslow, David C.
TITLE OF INVENTION: Cloning and Expression of Plasmidium
TITLE OF INVENTION: falciiparum Transmission-Blocking Target Antigen, Pfs23G
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323.1708
FILING DATE: 13-OCT-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/010.409
FILING DATE: 29-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Quine, Jonathan A.
REGISTRATION NUMBER: P-41,261
REFERENCE/DOCKET NUMBER: 015280-113100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

RESULT 11
 PCT-US92-06300-1
 Sequence 1, Application PC/TUS9206300
 GENERAL INFORMATION:
 APPLICANT: Hurwitz, David R
 APPLICANT: Nathan, Margaret
 APPLICANT: Shani, Moshe
 TITLE OF INVENTION: Transgenic Protein Production
 NUMBER OF SEQUENCES: 1
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Rhone-Poulenc Rorer, Inc.
 STREET: 500 Virginia Ave., Bldg. 3A
 CITY: Ft. Washington
 STATE: Pennsylvania
 COUNTRY: USA
 ZIP: 19034
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US92/06300
 FILING DATE: 19920730
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: Goodman, Rosanne
 REGISTRATION NUMBER: 52,534
 REFERENCE/DOCKET NUMBER: A0856-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 962-4130
 TELEFAX: (215) 962-4107
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 19557 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHEetical: NO
 ANTI-SENSE: NO
 PUBLICATION INFORMATION:
 AUTHORS: Minghetti, P P
 AUTHORS: Ruffner, D E
 AUTHORS: Kuang, W.-J.

[illegible]

DB 10757 AATATATATATATATATCCATCATATATAAATAATGAAATCTTAAATATATATACA 10816
 QY 169 AAGAGAGAGAGAGAGAAATACTACTT 195
 DB 10817 GAGATGAATTAACCTAATAGTAATT 10843

RESULT 13

US-08-449-315-36
 ; Sequence 36, Application US/08449315
 ; Patent No. 5650505

GENERAL INFORMATION:

APPLICANT: Ryals, John A.
 APPLICANT: Alexander, Danny C.
 APPLICANT: Beck, James J.
 APPLICANT: Duesing, John H.
 APPLICANT: Friedrich, Leslie B.
 APPLICANT: Goodman, Robert M.
 APPLICANT: Harms, Christian
 APPLICANT: Meins, Jr., Frederick
 APPLICANT: Montoya, Alice
 APPLICANT: Moyer, Mary B.
 APPLICANT: Neuhaus, Jean-Marc
 APPLICANT: Payne, George B.
 APPLICANT: Sperison, Christoph
 APPLICANT: Stinson, Jeffrey R.
 APPLICANT: Uknes, Scott J.
 APPLICANT: Ward, Eric R.
 APPLICANT: Williams, Shericea C.
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
 TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: New York
 COUNTRY: USA
 ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/449,315
 FILING DATE: 24-MAY-1995
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/161,271
 FILING DATE: 13-JAN-94
 APPLICATION NUMBER: US 08/093,301
 FILING DATE: 16-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/937,197
 FILING DATE: 6-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/678,378
 FILING DATE: 1-APR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/305,566
 FILING DATE: 6-FEB-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/165,667
 FILING DATE: 8-MAR-1988
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/042,847
 FILING DATE: 6-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/632,441
 FILING DATE: 21-DEC-1990
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/425,504
 FILING DATE: 20-OCT 1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/848,506
 FILING DATE: 6-MAR-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/768,122
 FILING DATE: 27-SEP-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/580,431
 FILING DATE: 7-SEP-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/368,672
 FILING DATE: 20-JUN-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/329,018
 FILING DATE: 24-MAR-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/045,957
 FILING DATE: 12-APR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmer, James Scott
 REGISTRATION NUMBER: 36,129
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919)541-8614
 TELEFAX: (919)541-8689
 INFORMATION FOR SEQ ID NO: 36:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 12124 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: Cucumis sativus
 INDIVIDUAL ISOLATE: Cucumber Chitinase Genomic DNA
 IMMEDIATE SOURCE:
 CLONE: pBscucchrht5
 US-08-449-315-36

Query Match 9.3%; Score 29.8; DB 2; Length 12124;
 Best Local Similarity 49.7%; Pred. No. 10;
 Matches 73; Conservative 1; Mismatches 73; Indels 0; Gaps 0;

QY 49 ACAACAAAAAGAAAAAGTTGTGACATGTAGTAGGAGTGTGACCCCTTACTCCCAT 108
 DB 10697 ACAGCAAAAAAATAATGCAAGGTTGAGTACCATTTGATGATGATATATAT 10756
 QY 109 CAAAAAATAATGATACATGTTAAAGCATARAAGGCAATATTTATCATATGTTCTA 168
 DB 10757 AATATATATATATATATTCATCATATATAAATAATGAAATCTTAAATATATATACA 10816
 QY 169 AAGAGAGAGAGAGAGAAATACTACTT 195
 DB 10817 GAGATGAATTAACCTAATAGTAATT 10843

RESULT 14

US-08-444-803-36
 ; Sequence 36, Application US/08444803
 ; Patent No. 5654414

GENERAL INFORMATION:

APPLICANT: Ryals, John A.
 APPLICANT: Alexander, Danny C.
 APPLICANT: Beck, James J.
 APPLICANT: Duesing, John H.
 APPLICANT: Friedrich, Leslie B.
 APPLICANT: Goodman, Robert M.
 APPLICANT: Harms, Christian
 APPLICANT: Meins, Jr., Frederick
 APPLICANT: Montoya, Alice

Query Match 9.3%; Score 29.8; DB 2; Length 12124;
Best Local Similarity 49.7%; Pred. No. 10;

Search completed: September 25, 1999, 07:30:22
Job time: 3459 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 12:05:37 ; Search time 1811.29 Seconds

(without alignments)
348,487 Million cell updates/sec

US-09-030-606-224

Title: Perfect score: 320
Sequence: 1 CCCCTGAGAGCTCTCTGTA.....TTTAACTCMGCATTGTGAC 320

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *
45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	202.2	63.2	512	21	R20590	R20590 YG49d06.r1
2	201.2	62.9	285	40	Z45397	Z45397 HSC2OD041.n
3	165.4	51.7	342	32	AA334754	AA334754 EST39036
4	152	47.5	177	25	N56292	N56292 JF5699F.Hum
5	136.8	42.8	315	32	AA373221	AA373221 EST85227
6	40.4	12.6	214	43	AU033864	AU033864 AU033864
7	38.4	12.0	700	48	AI557583	AI557583 PL2.1-1.D
8	37.2	11.6	310	40	C93124	C93124 C93124.DicT
9	36.8	11.5	249	41	AI054289	AI054289 q176C04.x
10	36.8	11.5	269	42	AI144030	AI144030 q163C08.x
11	36.6	11.4	496	23	HI7024	HI7024 ym39d10.s1
12	36.6	11.4	434	23	L47084	L47084 SCMRAP171.C
13	36.6	11.4	300	35	C57199	C57199 C57199.YuJ1
14	36.6	11.4	300	36	C57822	C57822 C57822.YuJ1
15	36	11.2	300	36	C59863	C59863 C59863.YuJ1
16	35.8	11.2	354	44	AU038221	AU038221 AU038221
17	35.2	11.0	504	35	AA550294	AA550294 1423m3.gm
18	35	10.9	587	50	AI667004	AI667004 fc24d11.y
19	34.8	10.9	471	46	AI404883	AI404883 GH24733.S
20	34.6	10.8	432	36	AA604491	AA604491 n073a09.s
21	34.4	10.8	305	37	AA721414	AA721414 n274b11.s
22	34.4	10.8	520	39	AA825604	AA825604 o057h06.s
23	34.4	10.8	312	43	AI239707	AI239707 qh43b02.x
24	34.4	10.8	568	44	AU039059	AU039059 AU039059
25	34.4	10.8	384	49	AI630952	AI630952 ty99g12.x
26	34.2	10.7	442	26	W47600	W47600 zc35d07.r1
27	34.2	10.7	323	34	C24351	C24351 C24351.DicT
28	34.2	10.7	171	40	C90553	C90553 C90553.DicT
29	34.2	10.7	412	43	AI178240	AI178240 EST221905
30	34.2	10.7	406	51	AI705892	AI705892 UT-R-AC1-
31	34	10.6	543	34	C24372	C24372 C24372.DicT
32	34	10.6	699	45	AU001859	AU001859 AU001859
33	34	10.6	612	47	AI525038	AI525038 promtra-5
34	34	10.6	373	48	AI582163	AI582163 tg63b09.x
35	34	10.6	477	53	HSM000181	AI035888 Homo sapi
36	33.8	10.6	464	20	Z32876	Z32876 HHEA41N.Alt
37	33.8	10.6	466	22	R22059	R22059 yb25d02.s1
38	33.8	10.6	583	25	N50519	N50519 yb89b04.s1
39	33.8	10.6	388	25	N97950	N97950 1711C3.Czap
40	33.8	10.6	383	25	N97969	N97969 1731C3.Czap
41	33.8	10.6	427	31	AA279549	AA279549 z68b509.r
42	33.8	10.6	565	46	AI407578	AI407578 EST35868
43	33.8	10.6	353	47	AI522468	AI522468 fd19b08.x
44	33.8	10.6	312	49	AV020572	AV020572 AV020572
45	33.6	10.5	602	34	AA532223	AA532223 CPEST.501

ALIGNMENTS

RESULT 1
R20590 R20590 512 bp mRNA EST 18-APR-1995
LOCUS YG49d06.r1 Soares Infant Brain INIB Homo sapiens cDNA clone
DEFINITION IMAGE:36244 5', mRNA sequence.
ACCESSION R20590
NID 9775371
VERSION R20590.1 GI:775371

KEYWORDS	EST.
SOURCE	human.
ORGANISM	
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 512)
AUTHORS	Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucada, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rikkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisansk, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
TITLE	The WashU-Merck EST Project
JOURNAL	Unpublished (1995)
COMMENT	
CONTACT	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 850L, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 321 Source: IMAGE Consortium, LINT This clone is available royalty-free through LINT; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Seq primer: M13Rpl High quality sequence stop: 321. Location/Qualifiers 1. 512
FEATURES	
SOURCE	

```

/organism="Homo sapiens"
/db_xref="GDB:408745"
/db_xref="taxon:9606"
/clone_image="36244"
/clone_lib="Soares Infant brain INTB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: latmid BA; Site: 1; Not
I; Site: 2; Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5';
AATCGAGAGAAATTCGCGCGCAGCAATTTTTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the latmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      141 a      93 c      111 g      138 t      9 others
ORIGIN

```

Query Match	63.2%	Score	202.2	DB	21	Length	512
Best Local Similarity	96.9%	Pred. No.	7.8e-38				
Matches	216	Conservative	1	Mismatches	4	Indels	2
						Gaps	1

Oy 1 CCCCGAAGGCGCTTCGTTGTAGAAAAATAGTACAGTTACACCAATAGGAAACAACAAAAAGA 60
 Db 76 CCCCGAAGGCGCTTCGTTGTAGAAAAATAGTACAGTTACACCAATAGGAAACAACAAAAAGA 135
 Oy 61 AAAAGTTGTGCATCTTGTAGTAGGAGAGTGTACCCTTACTCCCATCAAAAAAAAAT 120
 Db 136 AAAAGTTGTGCATCTTGTAGTAGGAGAGTGTACCCTTACTCCCATCAAAAAAAAAT 195
 Oy 121 GGATCATCGTTAAAGATAGAAAGGCGCAATATTTATCATATGTTCTAAAAAGAGAA 180
 Db 136 GGATCATCGTTAAAGATAGAAAGGCGCAATATTTATCATATGTTCTAAAAAGAGAGAA 235
 Oy 181 GAGAAATACTACTTCTTCRAAA--TGGAAGCCCTTAAAGGTG 221
 Db 236 GAGAAATACTACTTCTTCRAAAATGGGAACCCCTTAAAGGG 298

RESULT	2			
Z45397				
LOCUS	Z45397	285 bp	mRNA	EST
				01-MAY-1998

DEFINITION	HSC20d04l normalized infant brain cDNA Homo sapiens cDNA clone c-20d04, mRNA sequence.
ACCESSION	U53397
NID	0574610
VERSION	245397.1
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 285) Auffray,C., Behar,G., Bols,F., Bouchier,C., da Silva,C., Davines,M.D., Duprat,S., Houllagette,R., Jumeau,M.N., Lamy,B., Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y., Sebastiani,Karakchis,C. and Tessier,A.
TITLE	IMAG: molecular integration of the analysis of the human genome and its expression
JOURNAL	C.R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE	9527753.
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1407250.

CONTACT: Genethon
 Genexpress-genethon
 Genethon Centre de recherche sur le Genome Humain
 1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
 Tel.: 33169472800
 Fax: 33160778698
 Email: genexpress@genethon.fr
 Single read:
 Genexpress_library_idt: C;
 Genexpress_sequence_idt: y1c-2cd04;.

FEATURES
 Location/Qualifiers
 1..285

BASE COUNT	ORIGIN
101 a	2 others
44 c	86 t
52 g	

Query Match	62.9%	Score 201.2	DB 40	Length 285
Best Local Similarity	98.0%	Pred. No. 1.4e-37		
Matches 200, Conservative	2	Mismatches	2	Indels 0
				Gaps 0

Qy	1	CCCCGAGGCGCTTCTGTTAGAAAAATGTACAGTTACACCAATAGGAGACACAAAAAGA	60
Db	82	CCCCGAGGCGTCTCTGTTAGAAAAATGTACAGTTACACCAATAGGAGACACAAAAAGA	14141
Qy	61	AAAAGTTGTGACATTTGTAGTAGGAGAGTGTTACCCCTTACTCCCATCAAAAAAAAAT	120
Db	142	AAAAGTTGTGACATTTGTAGTAGGAGAGTGTTACCCCTTACTCCCATCAAAAAAAAAT	201
Qy	121	GGATCATGCTTAAAGATAGAGGCGCAATTTTATCATATGTTCTTAAAGAGAGGAA	180
Db	202	GGATCATGCTTAAAGATAGAGGCGCAATTTTATCATATGTTCTTAAAGAGAGGAA	261
Qy	181	GAGAAATACTACTTTCRCRAAT	204
Db	262	GAGAAATACTACTTTCRCRAAT	285

RESULT 3
AA334754 342 bp mRNA EST 21-APR-1997
LOCUS EST39036 Embryo, 9 week Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION AA334754
NID 91987018
VERSION AA334754.1 GI:1987018
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 342)
Adams M.D., Kerlavage A.R., Fleischmann R.D., Fuldner R.A.,
Bult C.J., Lee N.H., Kirkness E.F., Weinstock K.G., Gocayne J.D.,
White O., Sutton G., Blake J.A., Brandon R.C., Man Wal C.,
Clayton R.A., Cline T.R., Cotton M.D., Barle-Hughes J., Fine L.D.,
Fitzgerald L.M., Fitzhugh W.M., Fritchman J.L., Geoghegan N.S.,
Glodek A., Gnehm C.L., Hanna M.C., Hedblom E., Hinkle P.S., Jr.,
Kelley J.M., Kelley J.C., Liu L.-I., Marmaros S.M., Merrick J.M.,
Moreno-Palanges R.F., McDonald L.A., Nguyen D.T., Pelligrino S.M.,
Phillips C.A., Ryder S.E., Scott J.L., Seudek D.M., Shirley R.,
Small K.V., Spriggs T.A., Utterback T.R., Weidman J.F., Li Y.,
Bednarek D.P., Cao L., Cepeda M.A., Coleman T.A., Collins E.J.,
Dimke D., Feng D.-F., Ferris A., Fischer C., Hastings G.A.,
He W.W., Hu J.S., Greene J.M., Gruber J., Hudson P., Kim A.K.,
Kozak D.L., Kunsch C., Hungjun J., Li H., Weissner P.S., Olsen H.,
Raymond L., Wei X.F., Wang J., Xu C., Yu G.L., Rubin S.M.,
Dillon P.J., Fannon M.R., Rosen C.A., Haseltine W.A., Fields C.,
Fraser C.M. and Venter J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)

JOURNAL MEDLINE
96026280
On May 9, 1995 this sequence version replaced gi:802598.
Other ESTs: THC168716
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/cdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES
Location/Qualifiers
1..342
/organism="Homo sapiens"
/db_xref="ATCC (Inhost):136461"
/db_xref="taxon:9606"
/clone_lib="Embryo, 9 week"
/dev_stage="embryo, 9 wks"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 93 a 66 c 71 g 109 t 3 others
ORIGIN

Query Match 51.7%; Score 165.4; DB 32; Length 342;
Best Local Similarity 96.4%; Pred. No. 2.8e-29;
Matches 163; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

DB 152 TTTATCATATGTTCTTAAAGAGAGAGAGAAATCTACTCTTCRAAATGGAAGCC 211
|||||
DB 1 TTTATCATATGTTCTTAAAGAGAGAGAGAAATCTACTCTTCRAAATGGAAGCC 60
|||||
QY 212 CTAAAGGCTTTGATCTGAAGAGACAAAGTGGCCGTCATCTCTTTARAGTTG 271
|||||
DB 61 CTAAAGGCTTTGATCTGAAGAGACAAAGTGGCCGTCATCTCTTTARAGTTG 120
|||||
QY 272 CATGACTTGACACGCGTAAGTGTGAGTTTARACTCMGCATTGTGAC 320
|||||

DB 121 CATGACTTGACACGCGTAAGTGTGAGTTTARACTCMGCATTGTGAC 169
|||||

RESULT 4
N56292 177 bp mRNA EST 20-FEB-1996
LOCUS J35699F Human fetal heart, lambda ZAP Express Homo sapiens cDNA
DEFINITION clone J35699 5', mRNA sequence.
ACCESSION N56292
NID 91199140
VERSION N56292.1 GI:1199140
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 177)
Liew C.C.
CDNAs from fetal heart
Unpublished (1995)
On May 18, 1995 this sequence version replaced gi:811148.

JOURNAL MEDLINE
Contact: Liew CC
Department of Laboratory Medicine and Pathobiology
University of Toronto
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
Tel: 4169788758
Fax: 4169785650
Email: liewcc@utcc.utoronto.ca
Seq primer: TCACAGAAATCGCAGCAG.

FEATURES
Location/Qualifiers
1..177
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="11"
/clone_lib="Human fetal heart, lambda ZAP Express"
/clone_lib="Human fetal heart, lambda ZAP Express"
/lab_host="E. coli XL1-Blue"
/note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2:
XhoI; mRNA was purified from human fetal hearts (8-10
weeks). cDNA was synthesized using a XhoI-0.150 dt
adaptor-primer. EcoRI adaptors were ligated, followed by
digestion with XhoI, for directional cloning into
predigested lambda ZAP Express."

BASE COUNT 69 a 27 c 32 g 49 t
ORIGIN

Query Match 47.5%; Score 152; DB 25; Length 177;
Best Local Similarity 92.9%; Pred. No. 3.9e-26;
Matches 158; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

DB 1 CCCCTGAAGCTTTCTTTAGAAAATAGTACGTTACCAACCAATAGAGAACAAAGAGA 60
|||||
DB 8 CCCCTGAAGCTTTCTTTAGAAAATAGTACGTTACCAACCAATAGAGAACAAAGAGA 67
|||||
QY 61 AAAAGTTGTGACATTGATGAGAGAGTGTRACCCCTTATCCCATCAAAAAAAT 120
|||||
DB 68 AAAAGTTGTGACATTGATGAGAGAGTGTRACCCCTTATCCCATCAAAAAAAT 127
|||||
QY 121 GGATACATGTTAAAGAGATBAAGGCAATTTTATCATATGTTCTTAA 170
|||||
DB 128 GGATACATGTTAAAGAGATBAAGGCAATTTTATCATATGTTCTTAA 177
|||||

RESULT 5
AA373221 315 bp mRNA EST 21-APR-1997
LOCUS EST85227 HSC172 cells 1 Homo sapiens cDNA 5' end, mRNA sequence.
DEFINITION AA373221
ACCESSION AA373221
NID 92025541
VERSION AA373221.1 GI:2025541

KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
EST.	human.			
Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia				
Eutheria; Primates; Carnivora; Hominoidea; Homo.				
1 (bases 1 to 315)				
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Eulder, R.A.,				

TITLE	Initial assessment of human gene diversity and expression patterns
JOURNAL	based upon 83 million nucleotides of cDNA sequence
MEDLINE	Nature 377 (6547 Suppl.), 3-174 (1995)
COMMENT	96026280
	On May 18, 1995 this sequence version replaced g1:811155.

LOCUS	AU0033864	214 bp	EST	28-APR-1999
DEFINITION	AU0033864 Dictyostellium discoideum str. (H.Urushihara) Dictyostellium discoideum cDNA clone SLB558, mRNA sequence.			
ACCESSION	AU0033864			
VERSION	93799288			
NID	AU0033864.1	GI:379288		
KEYWORDS	EST.			
SOURCE	Dictyostellium discoideum.			
ORGANISM	Dictyostellium discoideum			
REFERENCE	Eukaryota: Dictyostellida; Dictyostellium.			
AUTHORS	1 (bases 1 to 214)			
	Morjo,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,H., Mitsu,B.N., Pi,M., Saito,T., Takemoto,K., Yasukawa,H., Williams,J., Maeda,M., Takeuchi,I., Ochiali,H. and Tanaka,Y.			
TITLE	The Dictyostellium developmental cDNA project: generation and analysis of expressed sequence tags from the first-finger stage of development			
JOURNAL	DNA Res. 5 (6), 335-340 (1998)			
MEDLINE	99156227			
COMMENT	On May 19, 1998 this sequence version replaced gi:2046208.			

FEATURES	Location/Qualifiers
source	1. 214

LOCUS	A1557583	700 bp	mRNA	SST	23-MAR-1999
DEFINITION	pc2.1-1.1.D10.r	tumor2	Homo sapiens	CDNA 3'	mRNA sequence.
ACCESSION	A1557583				
NID	94489946				
VERSION	A1557583.1	GI:4489946			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
	1 (bases 1 to 700)				
	Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jun				
	Yu,J. and Hood,L.				

..

1
 2
 3
 4

SOURCE	ORGANISM
Caenorhabditis elegans.	Caenorhabditis elegans.
Eukariyeta, Metazoa; Nematoda, Secernentea; Rhabdilita; Rhabdilitida;	Eukariyeta, Metazoa; Nematoda, Secernentea; Rhabdilita; Rhabdilitida;
Rhabdilita; Rhabdilitida; Rhabdilitidae; Peloderinae; Caenorhabditis	Rhabdilita; Rhabdilitida; Rhabdilitidae; Peloderinae; Caenorhabditis
1 (bases 1 to 300)	1 (bases 1 to 300)
Kohara, Y., Motolashvili, T., Tabara, H., Watanabe, H., Sugimoto, A.,	Kohara, Y., Motolashvili, T., Tabara, H., Watanabe, H., Sugimoto, A.,
Sanjo, M., Miyata, A. and Nishiyaki, A.	Sanjo, M., Miyata, A. and Nishiyaki, A.

Wed-Sep-29 14:27:52 1999

us-09-030-606-224_1.rst

This Page Blank (uspt